

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
1 July 2004 (01.07.2004)

PCT

(10) International Publication Number  
**WO 2004/055050 A2**

(51) International Patent Classification<sup>7</sup>: **C07K 14/47**

(21) International Application Number:  
PCT/IB2003/006434

(22) International Filing Date:  
10 December 2003 (10.12.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/432,699 10 December 2002 (10.12.2002) US  
60/485,027 3 July 2003 (03.07.2003) US

(71) Applicants (for all designated States except US): **EN-DOCUBE SAS** [FR/FR]; Prologue Biotech - BP 700, Rue Pierre et Marie Curie, F-31 319 Labège Cedex (FR). **CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS** [FR/FR]; 3, rue Michel Ange, F-75794 Paris Cedex 6 (FR).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **GIRARD, Jean-Philippe** [FR/FR]; 9 chemin de Vieux Moulin, F-31320 Rebique (FR). **AMALRIC, François** [FR/FR]; 26 rue Bessières, F-31500 Toulouse (FR). **ROUSSIGNE,**

**Myriam** [FR/FR]; 4 Route de Leran, F-09600 La Bastide sur l'Hers (FR). **CLOUAIRE, Thomas** [FR/FR]; 160 Grande Rue St-Michel, F-31400 Toulouse (FR).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

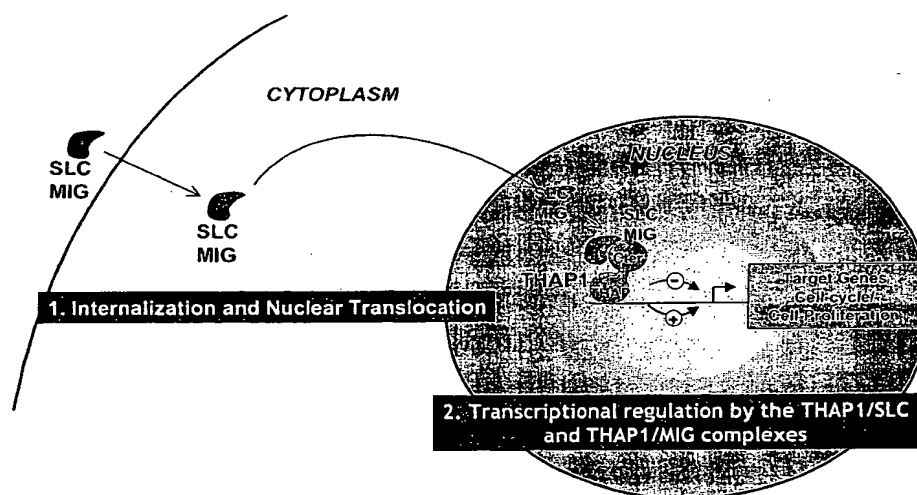
(84) Designated States (regional): ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: THAP PROTEINS AS NUCLEAR RECEPTORS FOR CHEMOKINES AND ROLES IN TRANSCRIPTIONAL REGULATION, CELL PROLIFERATION AND CELL DIFFERENTIATION



(57) Abstract: The invention relates to genes and proteins of the THAP family comprising a THAP domain, and their use in diagnostics, treatment of disease, and in the identification of molecules for the treatment of disease. The invention also relates to uses of THAP-type chemokine-binding agents, such as THAP-family proteins, as a nuclear receptors for a chemokines and to methods for the modulation (stimulation or inhibition) of transcription, cell proliferation and cell differentiation as well as methods for identifying for compounds which modulate THAP-chemokine interactions.

WO 2004/055050 A2

**THAP PROTEINS AS NUCLEAR RECEPTORS FOR CHEMOKINES  
AND ROLES IN TRANSCRIPTIONAL REGULATION,  
CELL PROLIFERATION AND CELL DIFFERENTIATION**

5

**FIELD OF THE INVENTION**

The present invention relates to genes and proteins of the THAP (THanatos (death)-Associated Protein) family, and uses thereof. In particular, the invention relates to the role of THAP-type chemokine-binding agents, such as THAP-family polypeptides, in transcriptional regulation and other chemokine-mediated cellular activities.

**BACKGROUND**

Coordination of cell proliferation and cell death is required for normal development and tissue homeostasis in multicellular organisms. A defect in the normal coordination of these two processes is a fundamental requirement for tumorigenesis.

Progression through the cell cycle is highly regulated, requiring the transit of numerous checkpoints (for review, see Hunter, 1993). The extent of cell death is physiologically controlled by activation of a programmed suicide pathway that results in morphologically recognizable form of death termed apoptosis (Jacobson et al, 1997; Vaux et al., 1994). Both extra-cellular signals, such as tumor necrosis factor, and intracellular signals, like p53, can induce apoptotic cell death. Although many proteins involved in apoptosis or the cell cycle have been identified, the mechanisms by which these two processes are coordinated are not well understood.

It is well established that molecules which modulate apoptosis have the potential to treat a wide range of conditions relating to cell death and cell proliferation. For example, such molecules may be used for inducing cell death for the treatment of cancers, inhibiting cell death for the treatment of neurodegenerative disorders, and inhibiting or inducing cell death for regulating angiogenesis. However, because many biological pathways controlling cell cycle and apoptosis have not yet been fully elucidated, there is a need for the identification of biological targets for the development of therapeutic molecules for the treatment of these disorders.

*PML nuclear bodies*

PML nuclear bodies (PML-NBs), also known as PODs (PML oncogenic domains), ND10 (nuclear domain 10) and Kr bodies, are discrete subnuclear domains that are specifically disrupted in cells from acute promyelocytic leukemia (APL), a distinct subtype of human myeloid leukemia (Maul et al., 2000 ; Ruggero et al., 2000 ; Zhong et al., 2000a). Their name derives from their most intensively studied protein component, the promyelocytic leukemia protein (PML), a RING finger IFN-inducible protein encoded by a gene originally cloned as the t(15 ;17) chromosomal translocation partner of the retinoic acid receptor (RAR) locus in APL. In APL cells, the presence of the leukemogenic fusion protein, PML-RAR, leads to the disruption of PML-NBs and the delocalization of PML and other PML-NB proteins into aberrant nuclear structures (Zhong et al.,

2000a). Treatment of both APL cell lines and patients with retinoic acid, which induces the degradation of the PML-RAR oncoprotein, results in relocation of PML and other NBs components into PML-NBs and complete remission of clinical disease, respectively. The deregulation of the PML-NBs by PML-RAR thus appears to play a critical role in tumorigenesis.

- 5 The analysis of mice, where the PML gene was disrupted by homologous recombination, has revealed that PML functions as a tumor suppressor *in vivo* (Wang et al., 1998a), that is essential for multiple apoptotic pathways (Wang et al., 1998b). Pml  $-/-$  mice and cells are protected from Fas, TNF $\alpha$ , ceramide and IFN-induced apoptosis as well as from DNA damage-induced apoptosis. However, the molecular mechanisms through which PML modulates the response to pro-apoptotic stimuli are not well understood (Wang et al., 1998b ; Quignon et al., 1998). Recent studies indicate that PML can participate in both p53-dependent and p53-independent apoptosis pathways (Guo et al., 2000 ; Fogal et al., 2000). p53-dependent DNA-damage induced apoptosis, transcriptional activation by p53 and induction of p53 target genes are all impaired in PML  $-/-$  primary cells (Guo et al., 2000). PML physically interacts with p53 and acts as a transcriptional co-activator for p53.
- 15 This co-activatory role of PML is absolutely dependent on its ability to recruit p53 in the PML-NBs (Guo et al., 2000; Fogal et al., 2000). The existence of a cross-talk between PML- and p53-dependent growth suppression pathways implies an important role for PML-NBs and PML-NBs-associated proteins as modulators of p53 functions. In addition to p53, the pro-apoptotic factor Daxx could be another important mediator of PML pro-apoptotic activities (Ishov et al., 1999; Zhong et al., 2000b; Li et al., 2000). Daxx was initially identified by its ability to enhance Fas-induced cell death. Daxx interacts with PML and localizes preferentially in the nucleus where it accumulates in the PML-NBs (Ishov et al., 1999; Zhong et al., 2000b; Li et al., 2000). Inactivation of PML results in delocalization of Daxx from PML-NBs and complete abrogation of Daxx pro-apoptotic activity (Zhong et al., 2000b). Daxx has recently been found to possess strong transcriptional repressor activity (Li et al., 2000). By recruiting Daxx to the PML-NBs, PML may inhibit Daxx-mediated transcriptional repression, thus allowing the expression of certain pro-apoptotic genes.

PML-NBs contain several other proteins in addition to Daxx and p53. These include the autoantigens Sp100 (Sternsdorf et al., 1999) and Sp100-related protein Sp140 (Bloch et al., 1999), the retinoblastoma tumor suppressor pRB (Alcalay et al., 1998), the transcriptional co-activator CBP (LaMorte et al., 1998), the Bloom syndrome DNA helicase BLM (Zhong et al., 1999) and the small ubiquitin-like modifier SUMO-1 (also known as sentrin-1 or PIC1; for recent reviews see Yeh et al., 2000; Melchior, 2000; Jentsch and Pyrowolakis, 2000). Covalent modification of PML by SUMO-1 (sumoylation) appears to play a critical role in PML accumulation into NBs (Muller et al., 1998) and the recruitment of other NBs components to PML-NBs (Ishov et al., 1999; Zhong et al., 2000c).

*Prostate apoptosis response-4*

Prostate apoptosis response-4 (PAR4) is a 38 kDa protein initially identified as the product of a gene specifically upregulated in prostate tumor cells undergoing apoptosis (for reviews see Rangnekar, 1998 ; Mattson et al., 1999). Consistent with an important role of PAR4 in apoptosis, induction of PAR4 in cultured cells is found exclusively during apoptosis and ectopic expression of PAR4 in NIH-3T3 cells (Diaz-Meco et al., 1996), neurons (Guo et al., 1998), prostate cancer and melanoma cells (Sells et al., 1997) has been shown to sensitize these cells to apoptotic stimuli. In addition, down regulation of PAR4 is critical for ras-induced survival and tumor progression (Barradas et al., 1999) and suppression of PAR4 production by antisense technology prevents apoptosis in several systems (Sells et al., 1997; Guo et al., 1998), including different models of neurodegenerative disorders (Mattson et al., 1999), further emphasizing the critical role of PAR4 in apoptosis. At the carboxy terminus, PAR4 contains both a leucine zipper domain (Par4LZ, amino acids 290-332), and a partially overlapping death domain (Par4DD, amino acids 258-332). Deletion of this carboxy-terminal part abrogates the pro-apoptotic function of PAR4 (Diaz-Meco et al., 1996 ; Sells et al., 1997 ; Guo et al., 1998). On the other hand, overexpression of PAR4 leucine zipper/death domain acts in a dominant negative manner to prevent apoptosis induced by full-length PAR4 (Sells et al., 1997 ; Guo et al., 1998). The PAR4 leucine zipper/death domain mediates PAR4 interaction with other proteins by recognizing two different kinds of motifs : zinc fingers of the Wilms tumor suppressor protein WT1 (Johnstone et al., 1996) and the atypical isoforms of protein kinase C (Diaz-Meco et al., 1996), and an arginine-rich domain from the death-associated-protein (DAP)-like kinase Dlk (Page et al., 1999). Among these interactions, the binding of PAR4 to aPKCs and the resulting inhibition of their enzymatic activity is of particular functional relevance because the aPKCs are known to play a key role in cell survival and their overexpression has been shown to abrogate the ability of PAR4 to induce apoptosis (Diaz-Meco et al., 1996 ; Berra et al., 1997).

## 25 *CHEMOKINES*

Chemokines (chemoattractant cytokines) are small secreted polypeptides of about 70-110 amino acids that regulate trafficking and effector functions of leukocytes, and play an important role in inflammation and host defense against pathogens (reviewed in Baggiolini M., et al. (1997) *Annu. Rev. immunol.* 15: 675-705; Proost P., et al. (1996) *Int. J. Clin. Lab. Res.* 26: 211-223; Premack, et al. (1996) *Nature Medicine* 2: 1174-1178; Yoshie, et al. (1997) *J. Leukocyte Biol.* 62: 634-644). Over 45 different human chemokines have been described to date. They vary in their specificities for different leukocyte types (neutrophils, monocytes, eosinophils, basophils, lymphocytes, dendritic cells, etc.), and in the types of cells and tissues where the chemokines are synthesized. Chemokines are typically produced at sites of tissue injury or stress, where they promote the infiltration of leukocytes into tissues and facilitate an inflammatory response. Some chemokines act selectively on immune system cells such as subsets of T-cells or B lymphocytes or antigen presenting cells, and may thereby promote immune responses to antigens. Some chemokines also



have the ability to regulate the growth or migration of hematopoietic progenitor and stem cells that normally differentiate into specific leukocyte types, thereby regulating leukocyte numbers in the blood.

The activities of chemokines are mediated by cell surface receptors which are members of the family of seven transmembrane, G-protein coupled receptors. At present, over fifteen different human chemokine receptors are known, including CCR1, CCR2, CCR3, CCR4, CCR5, CCR6, CCR7, CCR8, CCR9, CCR10, CCR11, CXCR1, CXCR2, CXCR3, CXCR4 and CXCR5. These receptors vary in their specificities for specific chemokines. Some receptors bind to a single known chemokine, while others bind to multiple chemokines. Binding of a chemokine to its receptor typically induces intracellular signaling responses such as a transient rise in cytosolic calcium concentration, followed by cellular biological responses such as chemotaxis.

Chemokine SLC/CCL21 (also known as SLC, CK $\beta$ -9, 6CKine, and exodus-2) is a member of the CC (beta)-chemokine subfamily, which shows 21 - 33% identity to other CC chemokines (Nagira, et al. (1997) J. Biol. Chem. 272:19518-19524; Hromas, et al. (1997) J. Immunol. 159:2554-2558; Hedrick, et al. (1997) J. Immunol. 159:1589-1593). SLC/CCL21 contains the four conserved cysteines characteristic of beta chemokines plus two additional cysteines in its unusually long carboxyl-terminal domain. Human SLC/CCL21 cDNA encodes a 134 amino acid residue, highly basic, precursor protein with a 23 amino acid residue signal peptide that is cleaved to form the predicted 111 amino acid residues mature protein. Mouse SLC/CCL21 cDNA encodes a 133 amino acid residue protein with 23 residue signal peptide that is cleaved to generate the 110 residue mature protein. Human and mouse SLC/CCL21 is highly conserved, exhibiting 86% amino acid sequence identity. The gene for human SLC/CCL21 has been localized at human chromosome 9p13 rather than chromosome 17, where the genes of many human CC chemokines are clustered. The SLC/CCL21 gene location is within a region of about 100 kb as the gene for MIP-3 beta/ELC/CCL19, another recently identified CC chemokine. SLC/CCL21 was previously known to be highly expressed in lymphoid tissues at the mRNA level, and to be a chemoattractant for T and B lymphocytes (Nagira, et al. (1997) J. Biol. Chem. 272:19518-19524; Hromas, et al. (1997) J. Immunol. 159:2554-2558; Hedrick, et al. (1997) J. Immunol. 159:1589-1593; Gunn, et al. (1998) Proc. Natl. Acad. Sci. 95:258-263). SLC/CCL21 also induces both adhesion of lymphocytes to intercellular adhesion molecule-1 and arrest of rolling cells (Campbell, et al. (1998) Science 279:381-384). All of the above properties are consistent with a role for SLC/CCL21 in regulating trafficking of lymphocytes through lymphoid tissues. Unlike most CC chemokines, SLC/CCL21 is not chemotactic for monocytes. However, it has been reported to inhibit hemopoietic progenitor colony formation in a dose-dependent manner (Hromas et al. (1997) J. Immunol. 159: 2554-58).

Chemokine SLC/CCL21 is a ligand for chemokine receptor CCR7 (Rossi et al. (1997) J. Immunol. 158:1033; Yoshida et al. (1997) J. Biol. Chem. 272:13803; Yoshida et al. (1998) J. Biol. Chem. 273:7118; Campbell et al. (1998) J Cell Biol 141:1053). CCR7 is expressed on T cells and

dendritic cells (DC), consistent with the chemotactic action of SLC/CCL21 for both lymphocytes and mature DC. Both memory (CD45RO<sup>+</sup>) and naïve (CD45RA<sup>+</sup>) CD4<sup>+</sup> and CD8<sup>+</sup> T cells express the CCR7 receptor (Sallusto et al. (1999) Nature 401:708). Within the memory T cell population, CCR7 expression discriminates between T cells with effector function that can migrate to inflamed tissues (CCR7<sup>+</sup>) vs. T cells that require a secondary stimulus prior to displaying effector functions (CCR7<sup>-</sup>) (Sallusto et al. (1999) Nature 401:708). Unlike mature DC, immature DC do not express CCR7 nor do they respond to the chemotactic action of CCL21 (Sallusto et al. (1998) Eur. J. Immunol. 28:2760; Dieu et al. (1998) J. Exp. Med. 188:373).

A key function of CCR7 and its two ligands SLC/CCL21 and MIP3b/CCL19 is facilitating recruitment and retention of cells to secondary lymphoid organs in order to promote efficient antigen exposure to T cells. CCR7-deficient mice demonstrate poorly developed secondary organs and exhibit an irregular distribution of lymphocytes within lymph nodes, Peyer's patches, and splenic periarteriolar lymphoid sheaths (Forster et al. (1999) Cell 99:23). These animals have severely impaired primary T cell responses largely due to the inability of interdigitating DC to migrate to the lymph nodes (Forster et al. (1999) Cell 99:23). The overall findings to date support the notion that CCR7 and its two ligands, CCL19 and CCL21, are key regulators of T cell responses via their control of T cell/DC interactions. CCR7 is an important regulatory molecule with an instructive role in determining the migration of cells to secondary lymphoid organs (Forster et al. (1999) Cell 99:23; Nakano et al. (1998) Blood 91:2886).

#### SUMMARY OF THE INVENTION

##### *THAP1 (THanatos-Associated-Protein-1)*

In the past few years, the inventors have focused on the molecular characterization of novel genes expressed in the specialized endothelial cells (HEVECs) of post-capillary high endothelial venules (Girard and Springer, 1995a; Girard and Springer, 1995b; Girard et al., 1999). In the present invention, they report the analysis of THAP1 (for THanatos (death)-Associated Protein-1), a protein that localizes to PML-NBs. Two hybrid screening of an HEVEC cDNA library with the THAP1 bait lead to the identification of a unique interacting partner, the pro-apoptotic protein PAR4. PAR4 is also found to accumulate into PML-NBs and targeting of the THAP1 / PAR4 complex to PML-NBs is mediated by PML. Similarly to PAR4, THAP1 is a pro-apoptotic polypeptide. Its pro-apoptotic activity requires a novel protein motif in the amino-terminal part called THAP domain. Together these results define a novel PML-NBs pathway for apoptosis that involves the THAP1/PAR4 pro-apoptotic complex.

Embodiments of the present invention includes genes, proteins and biological pathways involved in apoptosis. In some embodiments, the genes, proteins, and pathways disclosed herein may be used for the development of polypeptide, nucleic acid or small molecule therapeutics.

One embodiment of the present invention provides a novel protein motif, the THAP domain. The present inventors initially identified the THAP domain as a 90 residue protein motif in

the amino-terminal part of THAP1 and which is essential for THAP1 pro-apoptotic activity. THAP1 (THanatos (death) Associated Protein-1), as determined by the present inventors, is a pro-apoptotic polypeptide which forms a complex with the pro-apoptotic protein PAR4 and localizes in discrete subnuclear domains known as PML nuclear bodies. However, the THAP domain also  
5 defines a novel family of proteins, the THAP family, and the inventors have also provided at least twelve distinct members in the human genome (THAP-0 to THAP11), all of which contain a THAP domain (typically 80-90 amino acids) in their amino-terminal part. The present invention thus includes nucleic acid molecules, including in particular the complete cDNA sequences, encoding members of the THAP family, portions thereof encoding the THAP domain or polypeptides  
10 homologous thereto, as well as to polypeptides encoded by the THAP family genes. The invention thus also includes diagnostic and activity assays, and uses in therapeutics, for THAP family proteins or portions thereof, as well as drug screening assays for identifying compounds capable of inhibiting (or stimulating) pro-apoptotic activity of a THAP family member.

In one example of a THAP family member, THAP1 is determined to be an apoptosis  
15 inducing polypeptide expressed in human endothelial cells (HEVECs), providing characterization of the THAP sequences required for apoptosis activity in the THAP1 polypeptide. In further aspects, the invention is also directed to the interaction of THAP1 with the pro-apoptotic protein PAR4 and with PML-NBs, including methods of modulating THAP1 / PAR4 interactions for the treatment of disease. The invention also concerns interaction between PAR4 and PML-NBs,  
20 diagnostics for detection of said interaction (or localization) and modulation of said interactions for the treatment of disease.

Compounds which modulate interactions between a THAP family member and a THAP-family target molecule, a THAP domain or THAP-domain target molecule, or a PAR4 and a PML-NBs protein may be used in inhibiting (or stimulating) apoptosis of different cell types in various  
25 human diseases. For example, such compounds may be used to inhibit or stimulate apoptosis of endothelial cells in angiogenesis-dependent diseases including but not limited to cancer, cardiovascular diseases, inflammatory diseases, and to inhibit apoptosis of neurons in acute and chronic neurodegenerative disorders, including but not limited to Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral sclerosis, HIV encephalitis, stroke, epileptic seizures).

30 Oligonucleotide probes or primers hybridizing specifically with a THAP1 genomic DNA or cDNA sequence are also part of the present invention, as well as DNA amplification and detection methods using said primers and probes.

Fragments of THAP family members or THAP domains include fragments encoded by nucleic acids comprising at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200,  
35 500, or 1000 consecutive nucleotides selected from the group consisting of SEQ ID NOs: 160-175, or polypeptides comprising at least 8, 10, 12, 15, 20, 25, 30, 40, 50, 100, 150 or 200 consecutive amino acids selected from the group consisting of SEQ ID NOs: 1-114.

A further aspect of the invention includes recombinant vectors comprising any of the nucleic acid sequences described above, and in particular to recombinant vectors comprising a THAP1 regulatory sequence or a sequence encoding a THAP1 protein, THAP family member, THAP domain, fragments of THAP family members and THAP domains, homologues of THAP family members/ THAP domains, as well as to cell hosts and transgenic non human animals comprising said nucleic acid sequences or recombinant vectors.

Another aspect of the invention relates to methods for the screening of substances or molecules that inhibit or increase the expression of the THAP1 gene or genes encoding THAP family members, as well as with methods for the screening of substances or molecules that interact with and/or inhibit or increase the activity of a THAP1 polypeptide or THAP family polypeptide.

In accordance with another aspect, the present invention provides a medicament comprising an effective amount of a THAP family protein, e. g. THAP1, or a SLC/CCL21-binding fragment thereof, together with a pharmaceutically acceptable carrier. The medicaments described herein may be useful for treatment and/or prophylaxis.

As related to another aspect, the invention is concerned in particular with the use of a THAP family protein, homologs thereof and fragments thereof, for example THAP1, or a SLC/CCL21-binding fragment thereof as an anti-inflammatory agent. The THAP family protein, for example, THAP1 and fragments thereof will be useful for the treatment of conditions mediated by SLC/CCL21.

In a further aspect, the present invention provides a detection method comprising the steps of providing a SLC/CCL21 chemokine-binding molecule which is a THAP family protein, for example, THAP1, or an SLC/CCL21-binding fragment thereof, contacting the SLC/CCL21-binding THAP1 molecule with a sample, and detecting an interaction of the SLC/CCL21-binding THAP1 molecule with SLC/CCL21 chemokine in the sample.

In one example, the invention may be used to detect the presence of SLC/CCL21 chemokine in a biological sample. The SLC/CCL21-binding THAP1 molecule may be usefully immobilized on a solid support, for example as a THAP1/Fc fusion.

In accordance with another aspect, the present invention provides a method for inhibiting the activity of SLC/CCL21 chemokine in a sample, which method comprises contacting the sample with an effective amount of a SLC/CCL21 chemokine-binding molecule which is a THAP1 protein or a SLC/CCL21-binding fragment thereof.

In further aspects the invention provides a purified THAP1 protein or a SLC/CCL21-binding fragment thereof, or a THAP1/Fc fusion, for use in a method or a medicament as described herein; and a kit comprising such a purified THAP1 protein or fragment.

Some embodiments of the invention also envisage the use of fragments of the THAP1 protein, which fragments have SLC/CCL21 chemokine-binding properties. The fragments may be peptides derived from the protein. Use of such peptides can be preferable to the use of an entire

protein or a substantial part of a protein, for example because of the reduced immunogenicity of a peptide compared to a protein. Such peptides may be prepared by a variety of techniques including recombinant DNA techniques and synthetic chemical methods.

In addition to the above properties, THAP1 has the capability to bind to several additional  
5 chemokines. Such chemokines include, but are not limited to, ELC/CCL19, RANTES CCL5, MIG/CXCL9 and IP10/CXCL10. As such, further aspects of the present invention relate to the binding of chemokines by THAP1, a chemokine binding domain of THAP1, and polypeptides having at least 30% amino acid identity to THAP1 or a chemokine-binding domain of THAP1. Also contemplated is the binding of chemokines to oligomers and Fc immunoglobulin fusions of  
10 the above-listed polypeptides.

According to some aspects of the present invention, a THAP1 polypeptide, a chemokine-binding domain of THAP1, polypeptides having at least 30% amino acid identity to THAP1 or a chemokine-binding domain of THAP1 as well as oligomers or Fc immunoglobulin fusions of these  
15 proteins can be used in pharmaceutical compositions and/or medicaments for reducing the symptoms associated with inflammation and/or inflammatory diseases. As such, some aspects of the present invention include pharmaceutical compositions and/or medicaments comprising THAP1 protein, a chemokine-binding domain of THAP1, polypeptides having at least 30% amino acid identity to THAP1 or a chemokine-binding domain of THAP1 as well as oligomers or Fc immunoglobulin fusions of these proteins.

Yet other aspects of the invention relate THAP-family polypeptides, chemokine binding domains of THAP-family peptides, fusions of a THAP-family polypeptide with an immunoglobulin Fc region, fusions of a chemokine-binding domain of a THAP-family peptide with an immunoglobulin Fc region, oligomers of THAP family polypeptides, chemokine-binding domains of THAP family peptides, THAP-family peptide-Fc fusions, and chemokine-binding domain of  
25 THAP-family peptide-Fc fusions as well as polypeptides having at least 30% amino acid identity to any of the above-listed polypeptides. Pharmaceutical compositions which include one or more of these polypeptides are also contemplated.

Aspects of the invention relate to methods of binding a chemokine, inhibiting the activity of a chemokine, reducing or ameliorating the symptoms of a condition mediated or influenced by one or  
30 more chemokines, preventing the symptoms of a condition mediated or influenced by one or more chemokines and detecting a chemokine by using chemokine-binding agents such as THAP-family polypeptides, chemokine binding domains of THAP-family peptides, fusions of a THAP-family polypeptide with an immunoglobulin Fc region, fusions of a chemokine-binding domain of a THAP-family peptide with an immunoglobulin Fc region, oligomers of THAP family polypeptides, chemokine-binding domains of THAP family peptides, THAP-family peptide-Fc fusions, and  
35 chemokine-binding domain of THAP-family peptide-Fc fusions as well as polypeptides having at least 30% amino acid identity to any of the above-listed polypeptides.

Still other aspects of the present invention relate to methods modulating chemokine interactions with cellular receptors. Such receptors can be extracellular or can be molecules that are present within the cell. In some embodiments, chemokine interaction with one or more cellular receptors is modulated with one or more chemokine-binding agents, such as THAP-family polypeptides, chemokine binding domains of THAP-family peptides, fusions of a THAP-family polypeptide with an immunoglobulin Fc region, fusions of a chemokine-binding domain of a THAP-family peptide with an immunoglobulin Fc region, oligomers of THAP family polypeptides, chemokine-binding domains of THAP family peptides, THAP-family peptide-Fc fusions, and chemokine-binding domain of THAP-family peptide-Fc fusions as well as polypeptides having at least 30% amino acid identity to any of the above-listed polypeptides.

Some embodiments of the present invention relate to chemokines or chemokine complexes that are present within the nucleus of the cell and which modulate transcription. In some embodiments, complexes that are capable of modulating transcription comprise chemokines and chemokine-binding agents, such as THAP-family polypeptides, chemokine binding domains of THAP-family peptides, fusions of a THAP-family polypeptide with an immunoglobulin Fc region, fusions of a chemokine-binding domain of a THAP-family peptide with an immunoglobulin Fc region, oligomers of THAP family polypeptides, chemokine-binding domains of THAP family peptides, THAP-family peptide-Fc fusions, and chemokine-binding domain of THAP-family peptide-Fc fusions as well as polypeptides having at least 30% amino acid identity to any of the above-listed polypeptides. In some embodiments, the expression of one or more genes that are under the control of a THAP responsive promoter is modulated.

It will also be evident that the THAP-family proteins for use in the invention may be prepared in a variety of ways, in particular as recombinant proteins in a variety of expression systems. Any standard systems may be used such as baculovirus expression systems or mammalian cell line expression systems.

Other aspects of the invention are described in the following numbered paragraphs:

1. A method of identifying a candidate modulator of apoptosis comprising:
  - (a) contacting a THAP-family polypeptide or a biologically active fragment thereof with a test compound, wherein said THAP-family polypeptide comprises at least 30% amino acid identity to an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114; and
  - (b) determining whether said compound selectively modulates the activity of said polypeptide;wherein a determination that said test compound selectively modulates the activity of said polypeptide indicates that said compound is a candidate modulator of apoptosis.
2. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 3, or a biologically active fragment thereof.

3. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 4, or a biologically active fragment thereof.
4. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 5, or a biologically active fragment thereof.
- 5 5. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 6, or a biologically active fragment thereof.
6. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 7, or a biologically active fragment thereof.
7. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 8, or a biologically active fragment thereof.
- 10 8. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 9, or a biologically active fragment thereof.
9. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 10, or a biologically active fragment thereof.
- 15 10. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 11, or a biologically active fragment thereof.
11. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 12, or a biologically active fragment thereof.
12. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 13, or a biologically active fragment thereof.
- 20 13. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 14, or a biologically active fragment thereof.
14. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence selected from the group consisting of SEQ ID NOs: 15-114, and biologically active fragments thereof.
- 25 15. The method of Paragraph 1, wherein said biologically active fragment of said THAP-family protein has at least one biological activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis.
- 30 16. The methods of any one of Paragraphs 2-15 wherein said THAP-family polypeptide has at least one biological activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis.
- 35 17. An isolated nucleic acid encoding a polypeptide having apoptotic activity, said polypeptide consisting essentially of an amino acid sequence selected from the group consisting of :

- (a) amino acid positions 1-90 of SEQ ID NO: 2, a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto;
- (b) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 89 of SEQ ID NO: 3, a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto ;
- (c) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 89 of SEQ ID NO: 4, a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto ;
- (d) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 89 of SEQ ID NO: 5, a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto ;
- (e) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 90 of SEQ ID NO: 6, a fragment thereof having apoptotic activity or a polypeptide having at least 30% amino acid identity thereto ;
- (f) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 90 of SEQ ID NO: 7, a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto ;
- (g) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 90 of SEQ ID NO: 8, a fragment thereof having apoptotic activity ; or a polypeptide having at least 30% amino acid identity thereto ;
- (h) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 90 of SEQ ID NO: 9, a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto ;
- (i) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 92 of SEQ ID NO: 10, a fragment thereof having apoptotic activity or a polypeptide having at least 30% amino acid identity thereto ;
- (j) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 90 of SEQ ID NO: 11, a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto ;
- (k) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 90 of SEQ ID NO: 12, or a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto ;
- (l) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 90 of SEQ ID NO: 13, a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto ; and



(m) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 90 of SEQ ID NO: 14, a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto.

18. An isolated nucleic acid encoding a THAP-family polypeptide having apoptotic activity selected from the group consisting of:

(i) a nucleic acid molecule encoding a polypeptide comprising the amino acid sequence of a sequence selected from the group consisting of SEQ ID NOs: 1-114;

(ii) a nucleic acid molecule comprising the nucleic acid sequence of a sequence selected from the group consisting of SEQ ID NOs: 160-175 and the sequences complementary thereto; and

(iii) a nucleic acid the sequence of which is degenerate as a result of the genetic code to the sequence of a nucleic acid as defined in (i) and (ii).

19. The nucleic acid of Paragraph 18, wherein said nucleic acid comprises a nucleic acid selected from the group consisting of SEQ ID NOs. 5, 7, 8 and 11.

20. The nucleic acid of Paragraph 18, wherein said nucleic acid comprises a nucleic acid selected from the group consisting of SEQ ID NOs. 162, 164, 165 and 168.

21. An isolated nucleic acid encoding a THAP-family polypeptide having apoptotic activity comprising:

(i) the nucleic acid sequence of SEQ ID NOs : 1-2 or the sequence complementary thereto ; or

(ii) a nucleic acid molecule encoding a polypeptide comprising the amino acid sequence of SEQ ID NOs 1-2;

22. An isolated nucleic acid, said nucleic acid comprising a nucleotide sequence encoding:

i) a polypeptide comprising an amino acid sequence having at least about 80% identity to a sequence selected from the group consisting of the polypeptides of SEQ ID NOs: 1-114 and the polypeptides encoded by the nucleic acids of SEQ ID NOs: 160-175 or

ii) a fragment of said polypeptide which possesses apoptotic activity.

23. The nucleic acid of Paragraph of Paragraph 23, wherein said nucleic acid encodes a polypeptide comprising an amino acid sequence having at least about 80% identity to a sequence selected from the group consisting of the polypeptides of SEQ ID NOs: 5, 7, 8 and 11 and the polypeptides encoded by the nucleic acids of SEQ ID NOs: 162, 164, 165 and 168 or a fragment of said polypeptide which possesses apoptotic activity.

24. The nucleic acid of Paragraph 23, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of the sequences of SEQ ID NOs: 5, 7, 8 and 11 and the polypeptides encoded by the nucleic acids of SEQ ID NOs: 162, 164, 165 and 168.

25. The nucleic acid of Paragraph 23, wherein polypeptide identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters score=50 and

wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the default parameters of XBLAST.

26. The nucleic acid of Paragraph 17, wherein said nucleic acid is operably linked to a promoter.
- 5 27. An expression cassette comprising the nucleic acid of Paragraph 26.
28. A host cell comprising the expression cassette of Paragraph 27.
29. A method of making a THAP-family polypeptide, said method comprising providing a population of host cells comprising a recombinant nucleic acid encoding said THAP-family protein of any one of SEQ ID NOs. 1-114; and
- 10 culturing said population of host cells under conditions conducive to the expression of said recombinant nucleic acid;
- whereby said polypeptide is produced within said population of host cells.
30. The method of Paragraph 29 wherein said providing step comprises providing a population of host cells comprising a recombinant nucleic acid encoding said THAP-family protein
- 15 of any one of SEQ ID NOs. 5, 7, 8 and 11.
31. The method of Paragraph 29, further comprising purifying said polypeptide from said population of cells.
32. An isolated THAP polypeptide encoded by the nucleic acid of any one of SEQ ID Nos. 160-175.
- 20 33. The polypeptide of Paragraph 32, wherein said polypeptide is encoded by a nucleic acid selected from the group consisting of SEQ ID NOs. 5, 7, 8, 11, 162, 164, 165 and 168.
34. The polypeptide of Paragraph 32, wherein said polypeptide has at least one activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-
- 25 NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis.
35. An isolated THAP polypeptide or fragment thereof, said polypeptide comprising at least 12 contiguous amino acids of a sequence selected from the group consisting of SEQ ID NOs: 1-114.
- 30 36. The polypeptide of Paragraph 35, wherein said polypeptide comprises at least 12 contiguous amino acids of a sequence selected from the group consisting of SEQ ID NOs. 5, 7, 8, and 11.
37. The polypeptide of Paragraph 35, wherein said polypeptide has at least one activity selected from the group consisting of interaction with a THAP-family target protein, binding to a
- 35 nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis.

38. An isolated THAP polypeptide or fragment thereof, said polypeptide comprising an amino acid sequence having at least about 80% amino acid sequence identity to a sequence selected from the group consisting of SEQ ID NOs: 1-114 or a fragment thereof, said polypeptide or fragment thereof having at least one activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis.

39. The polypeptide of Paragraph 38, wherein said THAP polypeptide or fragment thereof comprises an amino acid sequence having at least about 80% amino acid sequence identity to a sequence selected from the group consisting of SEQ ID NOs: 5, 7, 8 and 11 or a fragment thereof having at least one activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis.

40. The polypeptide of Paragraph 38, wherein said polypeptide is selectively bound by an antibody raised against an antigenic polypeptide, or antigenic fragment thereof, said antigenic polypeptide comprising the polypeptide of any one of SEQ ID NOs: 1-114.

41. The polypeptide of Paragraph 38, wherein said polypeptide is selectively bound by an antibody raised against an antigenic polypeptide, or antigenic fragment thereof, said antigenic polypeptide comprising the polypeptide of any one of SEQ ID NOs: 5, 7, 8 and 11.

42. The polypeptide of Paragraph 38, wherein said polypeptide comprises the polypeptide of SEQ ID NOs: 1-114.

43. The polypeptide of Paragraph 38, wherein said polypeptide comprises a polypeptide selected from the group consisting of SEQ ID NOs: 5, 7, 8 and 11.

44. An antibody that selectively binds to the polypeptide of Paragraph 38.

45. An antibody according to Paragraph 44, wherein said antibody is capable of inhibiting binding of said polypeptide to a THAP-family target polypeptide.

46. An antibody according to Paragraph 44, wherein said antibody is capable of inhibiting apoptosis mediated by said polypeptide.

47. The polypeptide of Paragraph 38, wherein identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the default parameters of XBLAST.

48. A method of assessing the biological activity of a THAP-family polypeptide comprising:

- (a) providing a THAP-family polypeptide or a fragment thereof; and
- (b) assessing the ability of the THAP-family polypeptide to induce apoptosis of a cell.

49. A method of assessing the biological activity of a THAP-family polypeptide comprising:

- (a) providing a THAP-family polypeptide or a fragment thereof; and
- (b) assessing the DNA binding activity of the THAP-family polypeptide.

50. The method of Paragraphs 48 or 49, wherein step (a) comprises introducing to a cell a recombinant vector comprising a nucleic acid encoding a THAP-family polypeptide.

51. The method of Paragraphs 49 or 50, wherein the THAP-family polypeptide comprises a THAP consensus amino acid sequence depicted in SEQ ID NOs: 1-2, or a fragment thereof having at least one activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis.

52. The method of Paragraph 49, wherein the THAP-family polypeptide comprises an amino acid sequence selected from the group of sequences consisting of SEQ ID NOs: 1-114 or a fragment thereof having at least one activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis.

53. The method of Paragraph 49, wherein the THAP-family polypeptide comprises a native THAP-family polypeptide, or a fragment thereof having at least one activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis.

54. The method of Paragraph 49, wherein the THAP-family polypeptide comprises a THAP-family polypeptide or a fragment thereof having at least one activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis, wherein said THAP-family polypeptide or fragment thereof comprises at least one amino acid deletion, substitution or insertion.

55. An isolated THAP-family polypeptide comprising an amino acid sequence of SEQ ID NOs: 1-114, wherein said polypeptide comprises at least one amino acid deletion, substitution or insertion with respect to said amino acid sequence of SEQ ID NOs. 1-114.

56. A THAP-family polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114, wherein said polypeptide comprises at least one amino acid deletion, substitution or insertion with respect to said amino acid sequence of one of SEQ ID

NOs. 1-114 and displays a reduced ability to induce apoptosis or bind DNA compared to the wild-type polypeptide.

57. A THAP-family polypeptide comprising an amino acid sequence of SEQ ID NOs: 1-114, wherein said polypeptide comprises at least one amino acid deletion, substitution or insertion with respect to said amino acid sequence of one of SEQ ID NOs. 1-114 and displays a increased ability to induce apoptosis or bind DNA compared to the wild-type polypeptide.

58. A method of determining whether a THAP-family polypeptide is expressed within a biological sample, said method comprising the steps of :

(a) contacting a biological sample from a subject with:  
10 a polynucleotide that hybridizes under stringent conditions to a nucleic acid of SEQ ID NOs: 160-175 or

a detectable polypeptide that selectively binds to the polypeptide of SEQ ID NOs: 1-114; and

(b) detecting the presence or absence of hybridization between said polynucleotide and an RNA species within said sample, or the presence or absence of binding of said detectable polypeptide to a polypeptide within said sample;

wherein a detection of said hybridization or of said binding indicates that said THAP-family polypeptide is expressed within said sample.

59. The method of Paragraph 58, wherein said subject suffers from, is suspected of suffering from, or is susceptible to a cell proliferative disorder.

60. The method of Paragraph 59, wherein said cell proliferative disorder is a disorder related to regulation of apoptosis.

61. The method of Paragraph 58, wherein said polynucleotide is a primer, and wherein said hybridization is detected by detecting the presence of an amplification product comprising said primer sequence.

62. The method of Paragraph 58, wherein said detectable polypeptide is an antibody.

63. A method of assessing THAP-family activity in a biological sample, said method comprising the steps of :

(a) contacting a nucleic acid molecule comprising a binding site for a THAP-family polypeptide with :

30 (i) a biological sample from a subject or

(ii) a THAP-family polypeptide isolated from a biological sample from a subject, the polypeptide comprising the amino acid sequences of one of SEQ ID NOs: 1-114; and

(b) assessing the binding between said nucleic acid molecule and a THAP-family polypeptide wherein a detection of decreased binding compared to a reference THAP-family nucleic acid binding level indicates that said sample comprises a deficiency in THAP-family activity.

64. A method of determining whether a mammal has an elevated or reduced level of THAP-family expression, said method comprising the steps of :

(a) providing a biological sample from said mammal; and

(b) comparing the amount of a THAP-family polypeptide of SEQ ID NOs: 1-114 or of a THAP-family RNA species encoding a polypeptide of SEQ ID NOs: 1-114 within said biological sample with a level detected in or expected from a control sample ;

5 wherein an increased amount of said THAP-family polypeptide or said THAP-family RNA species within said biological sample compared to said level detected in or expected from said control sample indicates that said mammal has an elevated level of THAP-family expression, and wherein a decreased amount of said THAP-family polypeptide or said THAP-family RNA species within said biological sample compared to said level detected in or expected from said control  
10 sample indicates that said mammal has a reduced level of THAP-family expression.

65. The method of Paragraph 64, wherein said mammal suffers from, is suspected of suffering from, or is susceptible to a cell proliferative disorder.

66. A method of identifying a candidate inhibitor of a THAP-family polypeptide, a candidate inhibitor of apoptosis, or a candidate compound for the treatment of a cell proliferative  
15 disorder, said method comprising:

(a) contacting a THAP-family polypeptide according to SEQ ID NOs: 1-114 or a fragment comprising a contiguous span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114 with a test compound; and

(b) determining whether said compound selectively binds to said polypeptide;

20 wherein a determination that said compound selectively binds to said polypeptide indicates that said compound is a candidate inhibitor of a THAP-family polypeptide, a candidate inhibitor of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder.

67. A method of identifying a candidate inhibitor of apoptosis, a candidate compound for the treatment of a cell proliferative disorder, or a candidate inhibitor of a THAP-family  
25 polypeptide of SEQ ID NOs: 1-114 or a fragment comprising a contiguous span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114, said method comprising:

(a) contacting said THAP-family polypeptide with a test compound; and

(b) determining whether said compound selectively inhibits at least one biological activity  
30 selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis;

wherein a determination that said compound selectively inhibits said at least one biological activity  
35 of said polypeptide indicates that said compound is a candidate inhibitor of a THAP-family polypeptide, a candidate inhibitor of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder.

68. A method of identifying a candidate inhibitor of apoptosis, a candidate compound for the treatment of a cell proliferative disorder, or a candidate inhibitor of a THAP-family polypeptide of SEQ ID NOs: 1-114 or a fragment comprising a contiguous span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114, said method comprising:

(a) contacting a cell comprising said THAP-family polypeptide with a test compound; and

(b) determining whether said compound selectively inhibits at least one biological activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis;

wherein a determination that said compound selectively inhibits said at least one biological activity of said polypeptide indicates that said compound is a candidate inhibitor of a THAP-family polypeptide, a candidate inhibitor of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder.

69. The method of Paragraphs 67 or 68, wherein step (b) comprises assessing apoptotic activity, and wherein a determination that said compound inhibits apoptosis indicates that said compound is a candidate inhibitor of said THAP-family polypeptide.

70. The method of Paragraph 68 comprising introducing a nucleic acid comprising the nucleotide sequence encoding said THAP-family polypeptide according to any one of Paragraphs 32-43 into said cell.

71. A polynucleotide according to any one of Paragraphs 17- 25 attached to a solid support.

72. An array of polynucleotides comprising at least one polynucleotide according to Paragraph 71.

73. An array according to Paragraph 72, wherein said array is addressable.

74. A polynucleotide according to any one of Paragraphs 17 to 25 further comprising a label.

75. A method of identifying a candidate activator of a THAP-family polypeptide, said method comprising :

a) contacting a THAP-family polypeptide according to SEQ ID NOs: 1-114 or a fragment comprising a contiguous span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114 with a test compound; and

b) determining whether said compound selectively binds to said polypeptide;

wherein a determination that said compound selectively binds to said polypeptide indicates that said compound is a candidate activator of said polypeptide.

76. A method of identifying a candidate activator of a THAP-family polypeptide of SEQ ID NOs: 1-114 or a fragment comprising a contiguous span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114, said method comprising:

- (a) contacting said polypeptide with a test compound; and
- 5 (b) determining whether said compound selectively activates at least one biological activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis;

10 wherein a determination that said compound selectively activates said at least one biological activity of said polypeptide indicates that said compound is a candidate activator of said polypeptide.

77. A method of identifying a candidate activator of a THAP-family polypeptide of SEQ ID NOs: 1-114 or, a fragment comprising a contiguous span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114, said method comprising:

- (a) contacting a cell comprising said THAP-family polypeptide with a test compound; and
- 15 (b) determining whether said compound selectively activates at least one biological activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis;

wherein a determination that said compound selectively activates said at least one biological activity of said polypeptide indicates that said compound is a candidate activator of said polypeptide.

25 78. The method of Paragraphs 76 or 77, wherein said determining step comprises assessing apoptotic activity, and wherein a determination that said compound increases apoptosis activity indicates that said compound is a candidate activator of said THAP-family polypeptide.

79. The method of Paragraph 77 wherein step a) comprises introducing a nucleic acid comprising the nucleotide sequence encoding said THAP-family polypeptide according to any one of Paragraphs 17-25 into said cell.

30 80. A method of identifying a candidate modulator of PAR4 activity, said method comprising:

- (a) providing a PAR4 polypeptide or a fragment thereof; and
- (b) providing a PML-NB polypeptide, or a polypeptide associated with PML-NBs, or a  
35 fragment thereof; and
- (c) determining whether a test compound selectively modulates the ability of said PAR4 polypeptide to bind to said PML-NB polypeptide or polypeptide associated with PML-NBs;



wherein a determination that said test compound selectively inhibits the ability of said PAR4 polypeptide to bind to said PML-NB polypeptide or polypeptide associated with PML-NBs indicates that said compound is a candidate modulator of PAR4 activity.

81. A method of identifying a candidate modulator of PAR4 activity, said method  
5 comprising:

(a) providing a PAR4 polypeptide or a fragment thereof; and

(b) determining whether a test compound selectively modulates the ability of said PAR4 polypeptide to localise in PML-NBs;

wherein a determination that said test compound selectively inhibits the ability of said PAR4  
10 polypeptide to localise in PML-NBs indicates that said compound is a candidate modulator of PAR4 activity.

82. A method of identifying a candidate inhibitor of THAP-family activity, said method comprising:

(a) providing a THAP-family polypeptide of SEQ ID NOs: 1-114 or, a fragment comprising a  
15 contiguous span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114; and

(b) providing a THAP-family target polypeptide or a fragment thereof; and

(c) determining whether a test compound selectively inhibits the ability of said THAP-family polypeptide to bind to said THAP-family target polypeptide;

20 wherein a determination that said test compound selectively inhibits the ability of said THAP-family polypeptide to bind to said THAP-family target polypeptide indicates that said compound is a candidate inhibitor of THAP-family activity.

83. The method of Paragraph 82, comprising providing a cell comprising:

(a) a first expression vector comprising a nucleic acid encoding a THAP-family polypeptide of  
25 SEQ ID NOs: 1-114 or, a fragment comprising a contiguous span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114; and

(b) a second expression vector comprising a nucleic acid encoding a THAP-family target polypeptide, or a fragment thereof.

84. The method of Paragraph 82, wherein said THAP-family activity is apoptosis  
30 activity.

85. The method of Paragraph 82, wherein said THAP-family target protein is PAR-4.

86. The method of Paragraph 82, wherein said THAP-family polypeptide is a THAP-1, THAP-2 or THAP-3 protein and said THAP-family target protein is PAR-4.

87. A method of modulating apoptosis in a cell comprising modulating the activity of a  
35 THAP-family protein.

88. The method of Paragraph 87, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs. 1-114.

89. A method of modulating apoptosis in a cell comprising modulating the recruitment of PAR-4 to a PML nuclear body.

90. The method of Paragraph 89 wherein modulating the activity of a THAP-family protein comprises modulating the interaction of a THAP-family protein and a THAP-family target protein.

91. The method of Paragraph 89 wherein modulating the activity of a THAP-family protein comprises modulating the interaction of a THAP-family protein and a PAR4 protein.

92. The method of Paragraph 91 comprising modulation the interaction between a THAP-1, THAP-2, or THAP-3 protein and a PAR-4 protein.

93. A method of modulating the recruitment of PAR-4 to a PML nuclear body comprising modulating the interaction of said PAR-4 protein and a THAP-family protein.

94. The method of Paragraph 93, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs. 1-114.

95. A method of modulating angiogenesis in an individual comprising modulating the activity of a THAP-family protein in said individual.

96. The method of Paragraph 95, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs. 1-114.

97. A method of preventing cell death in an individual comprising inhibiting the activity of a THAP-family protein in said individual.

98. The method of Paragraph 97, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs. 1-114.

99. The method according to Paragraph 97, wherein the activity of said THAP-family protein is inhibited in the CNS.

100. A method of inducing angiogenesis in an individual comprising inhibiting the activity of a THAP-family protein in said individual.

101. The method of Paragraph 100, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs. 1-114.

102. A method according to Paragraph 100, wherein the activity of said THAP-family protein is inhibited in endothelial cells.

103. A method of inhibiting angiogenesis or treating cancer in an individual comprising increasing the activity of a THAP-family protein in said individual.

104. The method of Paragraph 103, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs. 1-114.

105. A method of treating inflammation or an inflammatory disorder in an individual comprising increasing the activity of a THAP-family protein in said individual.

106. The method of Paragraph 105, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs. 1-114.

107. A method according to Paragraphs 103 or 105, wherein the activity of said THAP-family protein is increased in endothelial cells.
108. A method of treating cancer in an individual comprising increasing the activity of a THAP-family protein in said individual.
- 5 109. The method of Paragraph 108, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs. 1-114.
110. The method of Paragraph 108, wherein increasing the activity of said THAP family protein induces apoptosis, inhibits cell division, inhibits metastatic potential, reduces tumor burden, increases sensitivity to chemotherapy or radiotherapy, kills a cancer cell, inhibits the growth of a cancer cell, kills an endothelial cell, inhibits the growth of an endothelial cell, inhibits angiogenesis, or induces tumor regression.
- 10 111. A method according to any one of Paragraphs 87-110, comprising contacting said subject with a recombinant vector encoding a THAP-family protein according to any one of Paragraphs 32-43 operably linked to a promoter that functions in said cell.
- 15 112. The method of Paragraph 111, wherein said promoter functions in an endothelial cell.
113. A viral composition comprising a recombinant viral vector encoding a THAP-family protein according to Paragraphs 32-43.
114. The composition of Paragraph 113, wherein said recombinant viral vector is an adenoviral, adeno-associated viral, retroviral, herpes viral, papilloma viral, or hepatitis B viral vector.
- 20 115. A method of obtaining a nucleic acid sequence which is recognized by a THAP-family polypeptide comprising contacting a pool of random nucleic acids with said THAP-family polypeptide or a portion thereof and isolating a complex comprising said THAP-family polypeptide and at least one nucleic acid from said pool.
- 25 116. The method of Paragraph 115 wherein said pool of nucleic acids are labeled.
117. The method of Paragraph 116 wherein said complex is isolated by performing a gel shift analysis.
118. A method of identifying a nucleic acid sequence which is recognized by a THAP-family polypeptide comprising:
- 30 (a) incubating a THAP-family polypeptide with a pool of labeled random nucleic acids;
- (b) isolating a complex between said THAP-family polypeptide and at least one nucleic acid from said pool;
- 35 (c) performing an amplification reaction to amplify the at least one nucleic acid present in said complex;

(d) incubating said at least one amplified nucleic acid with said THAP-family polypeptide;

(e) isolating a complex between said at least one amplified nucleic acid and said THAP-family polypeptide;

5 (f) repeating steps (c), (d) and (e) a plurality of times;

(g) determining the sequence of said nucleic acid in said complex.

119. A method of identifying a compound which inhibits the ability of a THAP-family polypeptide to bind to a nucleic acid comprising: incubating a THAP-family polypeptide or a fragment thereof which recognizes a binding site in a nucleic acid with a nucleic acid containing  
10 said binding site in the presence or absence of a test compound and determining whether the level of binding of said THAP-family polypeptide to said nucleic acid in the presence of said test compound is less than the level of binding in the absence of said test compound.

120. A method of identifying a test compound that modulates THAP-mediated activities comprising:

15 contacting a THAP-family polypeptide or a biologically active fragment thereof with a test compound, wherein said THAP-family polypeptide comprises an amino acid sequence having at least 30% amino acid identity to an amino acid sequence of SEQ ID NO: 1; and

20 determining whether said test compound selectively modulates the activity of said THAP-family polypeptide or biologically active fragment thereof, wherein a determination that said test compound selectively modulates the activity of said polypeptide indicates that said test compound is a candidate modulator of THAP-mediated activities.

121. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 1, or a biologically active fragment thereof.

25 122. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 2, or a biologically active fragment thereof.

123. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 3, or a biologically active fragment thereof.

30 124. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 4, or a biologically active fragment thereof.

125. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 5, or a biologically active fragment thereof.

126. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 6, or a biologically active fragment thereof.

35 127. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 7, or a biologically active fragment thereof.

128. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 8, or a biologically active fragment thereof.

129. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 9, or a biologically active fragment thereof.

5 130. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 10, or a biologically active fragment thereof.

131. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 11, or a biologically active fragment thereof.

10 132. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 12, or a biologically active fragment thereof.

133. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 13, or a biologically active fragment thereof.

134. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 14 or a biologically active fragments thereof.

15 135. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence selected from the group consisting of SEQ ID NOs: 15-114 or a biologically active fragments thereof.

136. The method of Paragraph 120, wherein said THAP-mediated activity is selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid, binding to PAR-4, binding to SLC, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis

137. The method of Paragraph 136, wherein said THAP-mediated activity is binding to PAR-4.

25 138. The method of Paragraph 136, wherein said THAP-mediated activity is binding to SLC.

139. The method of Paragraph 136, wherein said THAP-mediated activity is inducing apoptosis.

30 140. The method of Paragraph 136, wherein said nucleic acid comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 140-159.

141. The method of Paragraph 120, wherein said amino acid identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters, score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the default parameters of XBLAST.

35 142. An isolated or purified THAP domain polypeptide consisting essentially of an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-2, amino acids 1-89 of SEQ ID NOs: 3-5, amino acids 1-90 of SEQ ID NOs: 6-9, amino acids 1-92 of SEQ ID NO: 10,

amino acids 1-90 of SEQ ID NOs: 11-14 and homologs having at least 30% amino acid identity to any aforementioned sequence, wherein said polypeptide binds to a nucleic acid.

143. The isolated or purified THAP domain polypeptide of Paragraph 142 consisting essentially of SEQ ID NO: 1.

5 144. The isolated or purified THAP domain polypeptide of Paragraph 142, wherein said amino acid identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters, score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the default parameters of XBLAST.

10 145. The isolated or purified THAP domain polypeptide of Paragraph 142, wherein said nucleic acid comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 140-159.

146. An isolated or purified nucleic acid which encodes the THAP domain polypeptide of Paragraph 142 or a complement thereof.

15 147. An isolated or purified PAR4-binding domain polypeptide consisting essentially of an amino acid sequence selected from the group consisting of amino acids 143-192 of SEQ ID NO: 3, amino acids 132-181 of SEQ ID NO: 4, amino acids 186-234 of SEQ ID NO: 5, SEQ ID NO: 15 and homologs having at least 30% amino acid identity to any aforementioned sequence, wherein said polypeptide binds to PAR4.

20 148. The isolated or purified PAR4-binding domain of Paragraph 147 consisting essentially of SEQ ID NO: 15.

149. The isolated or purified PAR4-binding domain of Paragraph 147 consisting essentially of amino acids 143-193 of SEQ ID NO: 3.

150. The isolated or purified PAR4-binding domain of Paragraph 147 consisting essentially of amino acids 132-181 of SEQ ID NO: 4.

25 151. The isolated or purified PAR4-binding domain of Paragraph 147 consisting essentially of amino acids 186-234 of SEQ ID NO: 5.

30 152. The isolated or purified PAR4-binding domain polypeptide of Paragraph 147, wherein said amino acid identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters, score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the default parameters of XBLAST.

153. An isolated or purified nucleic acid which encodes the PAR4-binding domain polypeptide of Paragraph 147 or a complement thereof.

35 154. An isolated or purified SLC-binding domain polypeptide consisting essentially of an amino acid sequence selected from the group consisting of amino acids 143-213 of SEQ ID NO: 3 and homologs thereof having at least 30% amino acid identity, wherein said polypeptide binds to SLC.

155. The isolated or purified SLC-binding domain polypeptide of Paragraph 154, wherein said amino acid identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters, score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the default parameters of XBLAST.

5 156. An isolated or purified nucleic acid which encodes the SLC-binding domain polypeptide of Paragraph 154 or a complement thereof.

157. A fusion protein comprising an Fc region of an immunoglobulin fused to a polypeptide comprising an amino acid sequence selected from the group consisting of amino acids 143-213 of SEQ ID NO: 3 and homologs thereof having at least 30% amino acid identity.

10 158. An oligomeric THAP protein comprising a plurality of THAP polypeptides, wherein each THAP polypeptide comprises an amino acid sequence selected from the group consisting of amino acid 143-213 of SEQ ID NO: 3 and homologs thereof having at least 30% amino acid identity.

15 159. A medicament comprising an effective amount of a THAP1 polypeptide or an SLC-binding fragment thereof, together with a pharmaceutically acceptable carrier.

160. An isolated or purified THAP dimerization domain polypeptide consisting essentially of an amino acid sequence selected from the group consisting of amino acids 143 and 192 of SEQ ID NO: 3 and homologs thereof having at least 30% amino acid identity, wherein said polypeptide binds to a THAP-family polypeptide..

20 161. The isolated or purified THAP dimerization domain polypeptide of Paragraph 160, wherein said amino acid identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters, score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the default parameters of XBLAST.

25 162. An isolated or purified nucleic acid which encodes the THAP dimerization domain polypeptide of Paragraph 160 or a complement thereof.

163. An expression vector comprising a promoter operably linked to a nucleic acid having a nucleotide sequence selected from the group consisting of SEQ ID NOs: 160-175 and portions thereof comprising at least 18 consecutive nucleotides.

30 164. The expression vector of Paragraph 163, wherein said promoter is a promoter which is not operably linked to said nucleic acid selected from the group consisting of SEQ ID NOs.: 160-175 in a naturally occurring genome.

165. A host cell comprising the expression vector of Paragraph 163.

35 166. An expression vector comprising a promoter operably linked to a nucleic acid encoding a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114 and portions thereof comprising at least 18 consecutive nucleotides.

167. The expression vector of Paragraph 166, wherein said promoter is a promoter which is not operably linked to said nucleic acid selected from the group consisting of SEQ ID NOs: 160-175 in a naturally occurring genome.

168. A host cell comprising the expression vector of Paragraph 166.

169. A method of identifying a candidate inhibitor of a THAP-family polypeptide, a candidate inhibitor of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder, said method comprising:

contacting a THAP-family polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114 or a fragment comprising a span of at least 6 contiguous amino acids of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114 with a test compound; and

determining whether said compound selectively binds to said polypeptide, wherein a determination that said compound selectively binds to said polypeptide indicates that said compound is a candidate inhibitor of a THAP-family polypeptide, a candidate inhibitor of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder.

170. A method of identifying a candidate inhibitor of apoptosis, a candidate compound for the treatment of a cell proliferative disorder, or a candidate inhibitor of a THAP-family polypeptide of SEQ ID NOs: 1-114 or a fragment comprising a span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114, said method comprising:

contacting said THAP-family polypeptide with a test compound; and

determining whether said compound selectively inhibits at least one biological activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to SLC, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis, wherein a determination that said compound selectively inhibits said at least one biological activity of said polypeptide indicates that said compound is a candidate inhibitor of a THAP-family polypeptide, a candidate inhibitor of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder.

171. A method of identifying a candidate inhibitor of apoptosis, a candidate compound for the treatment of a cell proliferative disorder, or a candidate inhibitor of a THAP-family polypeptide of SEQ ID NOs: 1-114 or a fragment comprising a span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114, said method comprising:

contacting a cell comprising said THAP-family polypeptide with a test compound;

and

determining whether said compound selectively inhibits at least one biological activity selected from the group consisting of interaction with a THAP-family target



protein, binding to a nucleic acid sequence, binding to PAR-4, binding to SLC, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis, wherein a determination that said compound selectively inhibits said at least one biological activity of said polypeptide indicates that said compound is a candidate inhibitor of a THAP-family polypeptide, a candidate inhibitor of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder.

172. A method of identifying a candidate modulator of THAP-family activity, said method comprising:

providing a THAP-family polypeptide of SEQ ID NOs: 1-114 or, a fragment comprising a span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114; and

providing a THAP-family target polypeptide or a fragment thereof; and

determining whether a test compound selectively modulates the ability of said THAP-family polypeptide to bind to said THAP-family target polypeptide, wherein a determination that said test compound selectively modulates the ability of said THAP-family polypeptide to bind to said THAP-family target polypeptide indicates that said compound is a candidate modulator of THAP-family activity.

173. The method of Paragraph 172, wherein said THAP-family polypeptide is provided by a first expression vector comprising a nucleic acid encoding a THAP-family polypeptide of SEQ ID NOs: 1-114 or, a fragment comprising a contiguous span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114, and wherein said THAP-family target polypeptide is provided by a second expression vector comprising a nucleic acid encoding a THAP-family target polypeptide, or a fragment thereof.

174. The method of Paragraph 172, wherein said THAP-family activity is apoptosis activity.

175. The method of Paragraph 172, wherein said THAP-family target protein is PAR-4.

176. The method of Paragraph 172, wherein said THAP-family polypeptide is a THAP-1, THAP-2 or THAP-3 protein and said THAP-family target protein is PAR-4.

177. The method of Paragraph 172, wherein said THAP-family target protein is SLC.

178. A method of modulating apoptosis in a cell comprising modulating the activity of a THAP-family protein.

179. The method of Paragraph 178, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs: 1-114.

180. The method of Paragraph 178, wherein modulating the activity of a THAP-family protein comprises modulating the interaction of a THAP-family protein and a THAP-family target protein.

181. The method of Paragraph 178, wherein modulating the activity of a THAP-family protein comprises modulating the interaction of a THAP-family protein and a PAR4 protein.

182. A method of identifying a candidate activator of a THAP-family polypeptide, a candidate activator of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder, said method comprising:

contacting a THAP-family polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-98 or a fragment comprising a span of at least 6 contiguous amino acids of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-98 with a test compound; and

determining whether said compound selectively binds to said polypeptide, wherein a determination that said compound selectively binds to said polypeptide indicates that said compound is a candidate activator of a THAP-family polypeptide, a candidate activator of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder.

183. A method of identifying a candidate activator of apoptosis, a candidate compound for the treatment of a cell proliferative disorder, or a candidate activator of a THAP-family polypeptide of SEQ ID NOs: 1-98 or a fragment comprising a span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-98, said method comprising:

contacting said THAP-family polypeptide with a test compound; and

determining whether said compound selectively activates at least one biological activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to SLC, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis, wherein a determination that said compound selectively activates said at least one biological activity of said polypeptide indicates that said compound is a candidate activator of a THAP-family polypeptide, a candidate activator of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder.

184. A method of identifying a candidate activator of apoptosis, a candidate compound for the treatment of a cell proliferative disorder, or a candidate activator of a THAP-family polypeptide of SEQ ID NOs: 1 to 98 or a fragment comprising a span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-98, said method comprising:

contacting a cell comprising said THAP-family polypeptide with a test compound; and

determining whether said compound selectively activates at least one biological activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to SLC, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a

THAP-family target protein to PML-NBs, and inducing apoptosis, wherein a determination that said compound selectively activates said at least one biological activity of said polypeptide indicates that said compound is a candidate activator of a THAP-family polypeptide, a candidate activator of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder.

185. A method of ameliorating a condition associated with the activity of SLC in an individual comprising administering a polypeptide comprising the SLC binding domain of a THAP-family protein to said individual.

186. The method of Paragraph 185, wherein said polypeptide comprises a fusion protein comprising an Fc region of an immunoglobulin fused to a polypeptide comprising an amino acid sequence selected from the group consisting of amino acids 143-213 of SEQ ID NO: 3 and homologs thereof having at least 30% amino acid identity.

187. The method of Paragraph 185, wherein said polypeptide comprises an oligomeric THAP protein comprising a plurality of THAP polypeptides, wherein each THAP polypeptide comprises an amino acid sequence selected from the group consisting of amino acid 143-213 of SEQ ID NO: 3 and homologs thereof having at least 30% amino acid identity.

188. A method of modulating angiogenesis in an individual comprising modulating the activity of a THAP-family protein in said individual.

189. The method of Paragraph 188, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs: 1-114.

190. The method of Paragraph 188, wherein said modulation is inhibition.

191. The method of Paragraph 188, wherein said modulation is induction.

192. A method of reducing cell death in an individual comprising inhibiting the activity of a THAP-family protein in said individual.

193. The method of Paragraph 192, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs: 1-114.

194. The method according to Paragraph 192, wherein the activity of said THAP-family protein is inhibited in the CNS.

195. A method of reducing inflammation or an inflammatory disorder in an individual comprising modulating the activity of a THAP-family protein in said individual.

196. The method of Paragraph 195, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs: 1-114.

197. A method of reducing the extent of cancer in an individual comprising modulating the activity of a THAP-family protein in said individual.

198. The method of Paragraph 197, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs: 1-114.

199. The method of Paragraph 197, wherein increasing the activity of said THAP family protein induces apoptosis, inhibits cell division, inhibits metastatic potential, reduces tumor burden, increases sensitivity to chemotherapy or radiotherapy, kills a cancer cell, inhibits the growth of a cancer cell, kills an endothelial cell, inhibits the growth of an endothelial cell, inhibits angiogenesis,  
5 or induces tumor regression.

200. A method of forming a complex, said method comprising:  
contacting a chemokine with a chemokine-binding agent comprising a polypeptide selected from the group consisting of THAP-1, a polypeptide having at least 30% amino acid identity to THAP-1, a chemokine-binding domain of THAP-1 and a polypeptide  
10 having at least 30% amino acid identity to a chemokine-binding domain of THAP-1, wherein said chemokine and said chemokine binding agent form a complex.

201. The method of Paragraph 200, wherein said amino acid identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters, score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the  
15 default parameters of XBLAST.

202. The method of Paragraph 200, wherein said polypeptide is fused to an Fc region of an immunoglobulin.

203. The method of Paragraph 200, wherein said polypeptide comprises a THAP dimerization domain.

204. The method of Paragraph 203, wherein said THAP dimerization domain interacts with one or more THAP dimerization domains to form a THAP oligomer.

205. The method of Paragraph 200, wherein said polypeptide is a recombinant polypeptide.

206. The method of Paragraph 200, wherein said chemokine is selected from the group  
25 consisting of SLC, CCL19, CCL5, CXCL9 and CXCL10.

207. The method of Paragraph 200, wherein said chemokine is selected from the group consisting of SLC, CCL19 and CXCL9.

208. The method of Paragraph 200, wherein said polypeptide comprises THAP-1.

209. The method of Paragraph 208, wherein said THAP-1 comprises the amino acid  
30 sequence of SEQ ID NO: 3.

210. The method of Paragraph 200, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to THAP-1.

211. The method of Paragraph 200, wherein said polypeptide comprises a chemokine-binding domain of THAP-1.

212. The method of Paragraph 211, wherein said chemokine-binding domain of THAP-1  
35 comprises the amino acid sequence of amino acids 143-213 of SEQ ID NO: 3.

213. The method of Paragraph 200, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

214. A method of inhibiting the activity of a chemokine, said method comprising contacting a chemokine with an effective amount of an agent comprising a polypeptide selected from the group consisting of THAP-1, a polypeptide having at least 30% amino acid identity to THAP-1, a chemokine-binding domain of THAP-1 and a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1, wherein the activity of said chemokine is inhibited.

215. The method of Paragraph 214, wherein said amino acid identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters, score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the default parameters of XBLAST.

216. The method of Paragraph 214, wherein said polypeptide is fused to an Fc region of an immunoglobulin.

217. The method of Paragraph 214, wherein said polypeptide comprises a THAP dimerization domain.

218. The method of Paragraph 217, wherein said THAP dimerization domain interacts with one or more THAP dimerization domains to form a THAP oligomer.

219. The method of Paragraph 214, wherein said polypeptide is a recombinant polypeptide.

220. The method of Paragraph 214, wherein said polypeptide binds to a chemokine selected from the group consisting of SLC, CCL19, CCL5, CXCL9 and CXCL10.

221. The method of Paragraph 214, wherein said polypeptide binds to a chemokine selected from the group consisting of SLC, CCL19 and CXCL9.

222. The method of Paragraph 214, wherein said polypeptide comprises THAP-1.

223. The method of Paragraph 222, wherein said THAP-1 comprises the amino acid sequence of SEQ ID NO: 3.

224. The method of Paragraph 214, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to THAP-1.

225. The method of Paragraph 214, wherein said polypeptide comprises a chemokine-binding domain of THAP-1.

226. The method of Paragraph 225, wherein said chemokine-binding domain of THAP-1 comprises the amino acid sequence of amino acids 143-213 of SEQ ID NO: 3.

227. The method of Paragraph 214, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

228. A method of reducing inflammation comprising administering an effective amount of a chemokine binding agent to a subject afflicted with an inflammatory condition, wherein said

chemokine-binding agent comprises a polypeptide selected from the group consisting of THAP-1, a polypeptide having at least 30% amino acid identity to THAP-1, a chemokine-binding domain of THAP-1 and a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

5        229. The method of Paragraph 228, wherein said amino acid identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters, score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the default parameters of XBLAST.

10       230. The method of Paragraph 228, wherein said polypeptide is fused to an Fc region of an immunoglobulin.

231. The method of Paragraph 228, wherein said polypeptide comprises a THAP dimerization domain.

232. The method of Paragraph 231, wherein said THAP dimerization domain interacts with one or more THAP dimerization domains to form a THAP oligomer.

15       233. The method of Paragraph 228, wherein said polypeptide is a recombinant polypeptide.

234. The method of Paragraph 228, wherein said polypeptide binds to a chemokine selected from the group consisting of SLC, CCL19, CCL5, CXCL9 and CXCL10.

20       235. The method of Paragraph 228, wherein said polypeptide binds to a chemokine selected from the group consisting of SLC, CCL19 and CXCL9.

236. The method of Paragraph 228, wherein said polypeptide comprises THAP-1.

237. The method of Paragraph 236, wherein said THAP-1 comprises the amino acid sequence of SEQ ID NO: 3.

25       238. The method of Paragraph 228, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to THAP-1.

239. The method of Paragraph 228, wherein said polypeptide comprises a chemokine-binding domain of THAP-1.

240. The method of Paragraph 239, wherein said chemokine-binding domain of THAP-1 comprises the amino acid sequence of amino acids 143-213 of SEQ ID NO: 3.

30       241. The method of Paragraph 228, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

242. A method of reducing one or more symptoms associated with an inflammatory disease, said method comprising administering to a subject afflicted with said inflammatory disease a therapeutically effective amount of an agent which reduces or eliminates the activity of one or more chemokines, wherein said agent comprises a polypeptide selected from the group consisting of THAP-1, a polypeptide having at least 30% amino acid identity to THAP-1, a chemokine-

35

binding domain of THAP-1 and a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

243. The method of Paragraph 242, wherein said polypeptide is fused to an Fc region of an immunoglobulin.

5 244. The method of Paragraph 242, wherein said polypeptide comprises a THAP dimerization domain.

245. The method of Paragraph 244, wherein said THAP dimerization domain interacts with one or more THAP dimerization domains to form a THAP oligomer.

10 246. The method of Paragraph 242, wherein said polypeptide is a recombinant polypeptide.

247. The method of Paragraph 242, wherein said polypeptide binds to a chemokine selected from the group consisting of SLC, CCL19, CCL5, CXCL9 and CXCL10.

248. The method of Paragraph 242, wherein said polypeptide binds to a chemokine selected from the group consisting of SLC, CCL19 and CXCL9.

15 249. The method of Paragraph 242, wherein said polypeptide comprises THAP-1.

250. The method of Paragraph 249, wherein said THAP-1 comprises the amino acid sequence of SEQ ID NO: 3.

251. The method of Paragraph 242, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to THAP-1.

20 252. The method of Paragraph 242, wherein said polypeptide comprises a chemokine-binding domain of THAP-1.

253. The method of Paragraph 252, wherein said chemokine-binding domain of THAP-1 comprises the amino acid sequence of amino acids 143-213 of SEQ ID NO: 3.

25 254. The method of Paragraph 242, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

255. The method of Paragraph 242, wherein said inflammatory disease is arthritis.

256. The method of Paragraph 242, wherein said inflammatory disease is inflammatory bowel disease.

30 257. A method of detecting a chemokine, said method comprising:  
contacting a chemokine with a chemokine-binding agent comprising a polypeptide selected from the group consisting of THAP-1, a polypeptide having at least 30% amino acid identity to THAP-1, a chemokine-binding domain of THAP-1 and a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1; and  
detecting chemokine-binding agent bound to said chemokine.

35 258. The method of Paragraph 257, wherein chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL9 and CXCL10.

259. The method of Paragraph 257, wherein said chemokine is selected from the group consisting of SLC, CCL19 and CXCL9.

260. A detection system comprising a chemokine-binding agent comprising a polypeptide selected from the group consisting of THAP-1, a polypeptide having at least 30% amino acid identity to THAP-1, a chemokine-binding domain of THAP-1 and a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1, wherein said chemokine-binding agent is coupled to a solid support.

261. The detection system of Paragraph 260, wherein said polypeptide comprises THAP-1.

262. The detection system of Paragraph 261, wherein said THAP-1 comprises the amino acid sequence of SEQ ID NO: 3.

263. The detection system of Paragraph 260, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to THAP-1.

264. The detection system of Paragraph 260, wherein said polypeptide comprises a chemokine-binding domain of THAP-1.

265. The detection system of Paragraph 264, wherein said chemokine-binding domain of THAP-1 comprises the amino acid sequence of amino acids 143-213 of SEQ ID NO: 3.

266. The detection system of Paragraph 260, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

267. A pharmaceutical composition comprising a chemokine-binding agent in a pharmaceutically acceptable carrier, wherein said chemokine-binding agent comprises a polypeptide selected from the group consisting of THAP-1, a polypeptide having at least 30% amino acid identity to THAP-1, a chemokine-binding domain of THAP-1 and a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

268. The pharmaceutical composition of Paragraph 267, wherein said amino acid identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters, score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the default parameters of XBLAST.

269. The pharmaceutical composition of Paragraph 267, wherein said polypeptide is fused to an Fc region of an immunoglobulin.

270. The pharmaceutical composition of Paragraph 267, wherein said polypeptide comprises a THAP dimerization domain.

271. The pharmaceutical composition of Paragraph 271, wherein said THAP dimerization domain interacts with one or more THAP dimerization domains to form a THAP oligomer.

272. The pharmaceutical composition of Paragraph 267, wherein said polypeptide binds to a chemokine selected from the group consisting of SLC, CCL19, CCL5, CXCL9 and CXCL10.



273. The pharmaceutical composition of Paragraph 267, wherein said polypeptide binds to a chemokine selected from the group consisting of SLC, CCL19 and CXCL9.

274. The pharmaceutical composition of Paragraph 267, wherein said polypeptide comprises THAP-1.

5 275. The pharmaceutical composition of Paragraph 274, wherein said THAP-1 comprises the amino acid sequence of SEQ ID NO: 3.

276. The pharmaceutical composition of Paragraph 267, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to THAP-1.

10 277. The pharmaceutical composition of Paragraph 267, wherein said polypeptide comprises a chemokine-binding domain of THAP-1.

278. The pharmaceutical composition of Paragraph 277, wherein said chemokine-binding domain of THAP-1 comprises the amino acid sequence of amino acids 143-213 of SEQ ID NO: 3.

15 279. The pharmaceutical composition of Paragraph 267, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

280. A device for administering an agent, said device comprising a container that contains therein a chemokine-binding agent in a pharmaceutically acceptable carrier, wherein said chemokine-binding agent comprises a polypeptide selected from the group consisting of THAP-1, a polypeptide having at least 30% amino acid identity to THAP-1, a chemokine-binding domain of THAP-1 and a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

281. The device according to Paragraph 280, wherein said container is a syringe.

282. The device according to Paragraph 280, wherein said container is a patch for  
25 transdermal administration.

283. The device according to Paragraph 280, wherein said container is pressurized canister.

284. A kit comprising:

30 a chemokine-binding agent comprising a polypeptide selected from the group consisting of THAP-1, a polypeptide having at least 30% amino acid identity to THAP-1, a chemokine-binding domain of THAP-1 and a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1; and

instructions for using said chemokine-binding agent for detecting or inhibiting chemokines.

35 285. The kit of Paragraph 284, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL9 and CXCL10.

286. An isolated or purified chemokine-binding domain consisting essentially of a portion of SEQ ID NO: 3 that binds to a chemokine.

287. The isolated or purified chemokine-binding domain of Paragraph 286, wherein said chemokine is CCL19.

5 288. The isolated or purified chemokine-binding domain of Paragraph 286, wherein said chemokine is CCL5.

289. The isolated or purified chemokine-binding domain of Paragraph 286, wherein said chemokine is CXCL9.

10 290. The isolated or purified chemokine-binding domain of Paragraph 286, wherein said chemokine is CXCL10.

291. A method of modulating expression of a THAP responsive gene, said method comprising modulating the interaction of a THAP-family polypeptide or a biologically active fragment thereof with a nucleic acid, thereby enhancing or repressing expression of said THAP responsive gene.

15 292. The method of Paragraph 291, wherein said THAP-family polypeptide is THAP1.

293. The method of Paragraph 291, wherein said nucleic acid is a THAP responsive promoter.

294. The method of Paragraph 293, wherein said THAP responsive promoter comprises a THAP responsive element.

20 295. The method of Paragraph 294, wherein said THAP responsive element is a DR-5 element.

296. The method of Paragraph 294, wherein said THAP responsive element is an ER-11 element.

297. The method of Paragraph 294, wherein said THAP responsive element is THRE.

25 298. The method of Paragraph 293, wherein said THAP responsive promoter does not comprise a THAP responsive element.

299. The method of Paragraph 298, wherein said THAP responsive promoter is modulated by a product of a gene that is under the control of a promoter which comprises a THAP responsive element.

30 300. The method of Paragraph 291, wherein said THAP responsive gene is selected from the group consisting of Survivin, PTTG1/Securin, PTTG2/Securin, PTTG3/Securin, CKS1, MAD2L1, USP16/Ubp-M, HMMR/RHAMM, KIAA0008/HURP, CDCA7/JPO1 and THAP1.

301. The method of Paragraph 291, wherein said THAP responsive gene encodes a polypeptide involved in the G2 or M phase of the cell cycle.

35 302. The method of Paragraph 291, wherein said THAP responsive gene encodes a polypeptide involved in the S phase of the cell cycle.

303. The method of Paragraph 302, wherein said THAP responsive gene encodes a polypeptide involved in DNA replication.

304. The method of Paragraph 302, wherein said THAP responsive gene encodes a polypeptide involved in DNA repair.

5 305. The method of Paragraph 291, wherein said THAP responsive gene encodes a polypeptide involved in RNA splicing.

306. The method of Paragraph 291, wherein said THAP responsive gene encodes a polypeptide involved in apoptosis.

10 307. The method of Paragraph 291, wherein said THAP responsive gene encodes a polypeptide involved in angiogenesis.

308. The method of Paragraph 291, wherein said THAP responsive gene encodes a polypeptide involved in the proliferation of cancer cells.

309. The method of Paragraph 291, wherein said THAP responsive gene encodes a polypeptide involved in inflammatory disease.

15 310. A method of modulating the expression of a gene responsive to a THAP/chemokine complex, said method comprising modulating the interaction of a chemokine with a THAP-family polypeptide or a biologically active fragment thereof, thereby enhancing or repressing expression of said gene.

311. The method of Paragraph 310, wherein said THAP-family polypeptide is THAP1.

20 312. The method of Paragraph 310, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

313. The method of Paragraph 310, wherein said chemokine is SLC.

314. The method of Paragraph 310, wherein said chemokine is CXCL9.

25 315. The method of Paragraph 310, wherein the interaction between said chemokine and said THAP-family polypeptide is modulated by providing a THAP-type chemokine-binding agent.

316. The method of Paragraph 315, wherein said THAP-type chemokine-binding agent comprises a polypeptide selected from the group consisting of a THAP1 polypeptide, an chemokine-binding domain of a THAP1 polypeptide, a THAP1 polypeptide oligomer, an oligomer comprising a THAP1 chemokine-binding domain, a THAP1 polypeptide-immunoglobulin fusion, a 30 THAP1 chemokine-binding domain-immunoglobulin fusion and polypeptide homologs of any one of the aforementioned polypeptides.

317. The method of Paragraph 316, wherein said chemokine-binding domain is an SLC-binding domain.

35 318. The method of Paragraph 316, wherein said chemokine-binding domain is a CXCL9-binding domain.

319. The method of Paragraph 310, wherein said gene encodes a polypeptide involved in the G2 or M phase of the cell cycle.

320. The method of Paragraph 310, wherein said gene encodes a polypeptide involved in the S phase of the cell cycle.
321. The method of Paragraph 310, wherein said gene encodes a polypeptide involved in DNA replication.
- 5 322. The method of Paragraph 310, wherein said gene encodes a polypeptide involved in DNA repair.
323. The method of Paragraph 310, wherein said gene encodes a polypeptide involved in RNA splicing.
- 10 324. The method of Paragraph 310, wherein said gene encodes a polypeptide involved in apoptosis.
325. The method of Paragraph 310, wherein said gene encodes a polypeptide involved in angiogenesis.
326. The method of Paragraph 310, wherein said gene encodes a polypeptide involved in the proliferation of cancer cells.
- 15 327. The method of Paragraph 310, wherein said gene encodes a polypeptide involved in inflammatory disease.
328. A method of modulating the expression of a gene responsive to a THAP/chemokine complex, said method comprising modulating the interaction of a THAP/chemokine complex with a nucleic acid, thereby enhancing or repressing expression of said gene.
- 20 329. The method of Paragraph 328, wherein said THAP-family polypeptide is THAP1.
330. The method of Paragraph 328, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.
331. The method of Paragraph 328, wherein said chemokine is SLC.
332. The method of Paragraph 328, wherein said chemokine is CXCL9.
- 25 333. The method of Paragraph 328, wherein said gene encodes a polypeptide involved in the G2 or M phase of the cell cycle.
334. The method of Paragraph 328, wherein said gene encodes a polypeptide involved in the S phase of the cell cycle.
335. The method of Paragraph 334, wherein said gene encodes a polypeptide involved in DNA replication.
- 30 336. The method of Paragraph 334, wherein said gene encodes a polypeptide involved in DNA repair.
337. The method of Paragraph 328, wherein said gene encodes a polypeptide involved in RNA splicing.
- 35 338. The method of Paragraph 328, wherein said gene encodes a polypeptide involved in apoptosis.

339. The method of Paragraph 328, wherein said gene encodes a polypeptide involved in angiogenesis.
340. The method of Paragraph 328, wherein said gene encodes a polypeptide involved in the proliferation of cancer cells.
- 5 341. The method of Paragraph 328, wherein said gene encodes a polypeptide involved in inflammatory disease.
342. The method of Paragraph 328, wherein said nucleic acid is a THAP responsive promoter.
343. The method of Paragraph 342, wherein said THAP responsive promoter comprises  
10 a THAP responsive element.
344. The method of Paragraph 343, wherein said THAP responsive element is a DR-5 element.
345. The method of Paragraph 343, wherein said THAP responsive element is an ER-11 element.
- 15 346. The method of Paragraph 343, wherein said THAP responsive element is THRE.
347. The method of Paragraph 342, wherein said THAP responsive promoter does not comprise a THAP responsive element.
348. The method of Paragraph 347, wherein said THAP responsive promoter is modulated by a product of a gene that is under the control of a promoter which comprises a THAP  
20 responsive element.
349. A pharmaceutical composition comprising a THAP responsive element in a pharmaceutically acceptable carrier.
350. The pharmaceutical composition of Paragraph 349, wherein said THAP responsive element is a DR-5 element.
- 25 351. The pharmaceutical composition of Paragraph 349, wherein said THAP responsive element is an ER-11 element.
352. The pharmaceutical composition of Paragraph 349, wherein said THAP responsive element is an THRE.
353. A transcription factor decoy consisting essentially of a THAP responsive element.
- 30 354. The transcription factor decoy of Paragraph 353, wherein said THAP responsive element is a DR-5 element.
355. The transcription factor decoy of Paragraph 353, wherein said THAP responsive element is a ER-11 element.
356. The transcription factor decoy of Paragraph 353, wherein said THAP responsive  
35 element is a THRE element.
357. A cell comprising a transcription factor decoy of Paragraph 353.

358. A method of modulating the interaction between a nucleic acid and a THAP-family polypeptide or a biologically active fragment thereof, said method comprising providing a transcription factor decoy which comprises a THAP responsive element, thereby modulating the interaction between said nucleic acid and said THAP-family polypeptide or a biologically active  
5 fragment thereof.

359. The method of Paragraph 358, wherein said THAP-family polypeptide is THAP1.

360. The method of Paragraph 358, wherein said THAP responsive element is a DR-5 element.

361. The method of Paragraph 358, wherein said THAP responsive element is an ER-11  
10 element.

362. The method of Paragraph 358, wherein said THAP responsive element is THRE.

363. A method of modulating the interaction between a nucleic acid and a THAP/chemokine complex, said method comprising providing a transcription factor decoy which comprises a THAP responsive element, thereby modulating the interaction between said nucleic  
15 acid and said THAP/chemokine complex.

364. The method of Paragraph 363, wherein said THAP-family polypeptide is THAP1.

365. The method of Paragraph 363, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

366. The method of Paragraph 363, wherein said chemokine is SLC.

20 367. The method of Paragraph 363, wherein said chemokine is CXCL9.

368. The method of Paragraph 363, wherein said THAP responsive element is a DR-5 element.

369. The method of Paragraph 363, wherein said THAP responsive element is an ER-11  
25 element.

370. The method of Paragraph 363, wherein said THAP responsive element is THRE.

371. A vector packaging cell line comprising a cell comprising a viral vector which comprises a promoter operably linked to a nucleic acid encoding a THAP-family polypeptide or a biologically active fragment thereof.

372. The cell line of Paragraph 371, wherein said cell further comprises an introduced  
30 nucleic acid construct comprising a nucleic acid encoding a chemokine operably linked to a promoter.

373. The cell line of Paragraph 372, wherein said chemokine-encoding construct is included on the same vector as said nucleic acid encoding said THAP-family polypeptide or biologically active fragment thereof.

35 374. The cell line of Paragraph 372, wherein said nucleic acid encoding said chemokine encodes a chemokine selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

375. The cell line of Paragraph 372, wherein said nucleic acid encoding said chemokine encodes SLC.
376. The cell line of Paragraph 372, wherein said nucleic acid encoding said chemokine encodes CXCL9.
- 5 377. The cell line of Paragraph 371, wherein said THAP-family polypeptide is THAP1.
378. The cell line of Paragraph 371, wherein said cell is a mammalian cell.
379. The cell line of Paragraph 378, wherein said cell is a human cell.
380. The cell line of Paragraph 371, wherein said viral vector is an adenoviral vector.
381. The cell line of Paragraph 371, wherein said viral vector is a retroviral vector.
- 10 382. A cell which is genetically engineered to express a THAP-family polypeptide or a biologically active fragment thereof.
383. The cell line of Paragraph 382, wherein said THAP-family polypeptide is THAP1.
384. The cell line of Paragraph 382, wherein said cell is a mammalian cell.
385. The cell line of Paragraph 382, wherein said cell is a human cell.
- 15 386. The cell line of Paragraph 382, wherein said THAP family polypeptide is encoded by a gene that is introduced into the cell on an adenoviral vector.
387. The cell line of Paragraph 382, wherein said THAP family polypeptide is encoded by a gene that is introduced into the cell on a retroviral vector.
388. A method of constructing a cell which expresses a recombinant THAP-family polypeptide, said method comprising introducing into a cell a vector comprising a nucleic acid encoding a THAP-family polypeptide or a biologically active fragment thereof operably linked to a promoter.
- 20 389. The method of Paragraph 388, further comprising introducing into a cell a nucleic acid construct comprising a nucleic acid encoding a chemokine operably linked to a promoter.
- 25 390. The method of Paragraph 389, wherein said chemokine-encoding construct is included on the same vector as said nucleic acid encoding said THAP-family polypeptide or biologically active fragment thereof.
391. The method of Paragraph 389, wherein said nucleic acid encoding said chemokine encodes a chemokine selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.
- 30 392. The method of Paragraph 389, wherein said nucleic acid encoding said chemokine encodes SLC.
393. The method of Paragraph 389, wherein said nucleic acid encoding said chemokine encodes CXCL9.
- 35 394. The method of Paragraph 388, wherein said THAP-family polypeptide is THAP1.
395. The method of Paragraph 388, wherein said cell is a mammalian cell.
396. The method of Paragraph 395, wherein said cell is a human cell.

397. The method of Paragraph 388, wherein said vector is a viral vector.
398. The method of Paragraph 397, wherein said vector is an adenoviral vector.
399. The method of Paragraph 397, wherein said vector is a retroviral vector.
400. The method of Paragraph 388, wherein said vector is introduced into said cell by  
5 transfection.
401. A method of ameliorating symptoms associated with a condition mediated by a THAP/chemokine complex, said method comprising:  
introducing into a cell a nucleic acid construct comprising a nucleic acid encoding a  
chemokine operably linked to a promoter and a nucleic acid construct comprising a nucleic  
10 acid encoding a THAP-family polypeptide or a biologically active fragment thereof  
operably linked to a promoter; and  
expressing said nucleic acid encoding said chemokine and said nucleic acid  
encoding said THAP-family polypeptide or biologically active fragment thereof.
402. The method of Paragraph 401, wherein said nucleic acid constructs are present on a  
15 single vector.
403. The method of Paragraph 401, wherein said nucleic acid constructs are present on  
different vectors.
404. The method of Paragraph 401, wherein said cell is a mammalian cell.
405. The method of Paragraph 404, wherein said cell is a human cell.
- 20 406. The method of Paragraph 401, wherein said nucleic acid encoding said chemokine  
encodes a chemokine selected from the group consisting of SLC, CCL19, CCL5, CXCL11,  
CXCL10 and CXCL9.
407. The method of Paragraph 401, wherein said nucleic acid encoding said chemokine  
encodes SLC.
- 25 408. The method of Paragraph 401, wherein said nucleic acid encoding said chemokine  
encodes CXCL9.
409. The method of Paragraph 401, wherein said THAP-family polypeptide is THAP1.
410. A method of identifying a test compound that modulates transcription at a THAP  
responsive element, said method comprising:  
30 comparing the level of transcription from a THAP responsive promoter in the  
presence and absence of a test compound wherein a determination that the level of  
transcription is increased or decreased in the presence of said test compound relative to the  
level of transcription in the absence of said test compound indicates that said test compound  
is a candidate modulator of transcription.
- 35 411. The method of Paragraph 410, wherein the level of transcription from said THAP  
responsive promoter in the presence and absence of the test compound is determined by performing  
an in vitro transcription reaction using a construct comprising said THAP responsive promoter and



a THAP-family polypeptide or a biologically active fragment thereof, wherein said THAP-family polypeptide comprises an amino acid sequence having at least 30% amino acid identity to an amino acid sequence of SEQ ID NO: 1.

412. The method of Paragraph 410, wherein the level of transcription from said THAP responsive promoter in the presence and the absence of the test compound is determined by measuring the level of transcription from a THAP responsive promoter in a cell expressing a THAP-family polypeptide or a biologically active fragment thereof, wherein said THAP-family polypeptide comprises an amino acid sequence having at least 30% amino acid identity to an amino acid sequence of SEQ ID NO: 1.

413. The method of Paragraph 410, wherein said THAP-family polypeptide or biologically active fragment thereof is selected from the group consisting of SEQ ID NOs: 1-114 and biologically active fragments thereof.

414. The method of Paragraph 410, wherein said THAP responsive promoter comprises a THAP responsive element having a nucleotide sequence selected from the group consisting of SEQ ID NOs: 140-159, SEQ ID NO: 306, and homologs thereof having at least 60% nucleotide identity.

415. The method of Paragraph 411 or Paragraph 122, wherein the level of transcription in the presence or absence of said test compound is measured in the presence of a chemokine.

416. The method of Paragraph 415, wherein said chemokine is selected from the group consisting of CCL family chemokines and CXCL family chemokines.

417. The method of Paragraph 416, wherein said CCL family chemokine is selected from the group consisting of SLC, CCL19 and CCL5.

418. The method of Paragraph 416, wherein said CXCL family chemokine is selected from the group consisting of CXCL11, CXCL10 and CXCL9.

419. The method of Paragraph 415, wherein the level of transcription in the presence or absence of said test compound is measured in a cell which expresses a receptor for said chemokine.

420. The method of Paragraph 419, wherein said chemokine receptor is selected from the group consisting of CCR1, CCR3, CCR5, CCR7, CCR11 and CXCR3.

421. The method of Paragraph 420, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5; CXCL11, CXCL10 and CXCL9.

422. The method of Paragraph 419, wherein said THAP-family polypeptide comprises THAP1 or a biologically active fragment thereof and said cell expresses the CCR7 receptor.

423. The method of Paragraph 422, wherein said chemokine is SLC.

424. The method of Paragraph 419, wherein said THAP-family polypeptide comprises THAP1 or a biologically active fragment thereof and said cell expresses the CXCR3 receptor.

425. The method of Paragraph 424, wherein said chemokine is CXCL9.

426. The method of Paragraph 412, wherein said THAP responsive promoter is in a gene endogenous to said cell.

427. The method of Paragraph 412, wherein said THAP responsive promoter has been introduced into said cell.

5 428. The method of Paragraph 412, wherein said THAP responsive promoter does not comprise a THAP responsive element.

429. The method of Paragraph 428, wherein said THAP responsive promoter is modulated by a product of a gene that is under the control of a promoter which comprises a THAP responsive element.

10 430. A method for reducing the symptoms associated with a condition selected from the group consisting of excessive or insufficient angiogenesis, inflammation, metastasis of a cancerous tissue, excessive or insufficient apoptosis, cardiovascular disease and neurodegenerative diseases comprising modulating the interaction between a THAP-family polypeptide and a chemokine in an individual suffering from said condition.

15 431. The method of Paragraph 430, wherein said THAP-family polypeptide is selected from a group consisting of polypeptides having an amino acid sequence of SEQ ID NOs: 1-114.

432. The method of Paragraph 430, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

20 433. The method of Paragraph 430, wherein said chemokine is SLC and the condition is inflammation.

434. The method of Paragraph 430, wherein said chemokine is SLC and the condition is excessive or insufficient angiogenesis.

435. The method of Paragraph 430, wherein said chemokine is CXCL9 and the condition is inflammation.

25 436. The method of Paragraph 430, wherein said chemokine is CXCL9 and the condition is excessive or insufficient angiogenesis.

437. A method for reducing the symptoms associated with a condition resulting from the activity of a chemokine in an individual comprising modulating the interaction between said chemokine and a THAP-family polypeptide in said individual.

30 438. The method of Paragraph 437, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

439. The method of Paragraph 437, wherein said chemokine is SLC.

440. The method of Paragraph 437, wherein said chemokine is CXCL9.

441. The method of Paragraph 437, wherein said THAP-family polypeptide is THAP-1.

35 442. The method of Paragraph 437, wherein the condition is inflammation.

443. The method of Paragraph 437, wherein the condition is excessive or insufficient angiogenesis.

444. The method of Paragraph 437, wherein the interaction between said chemokine and said THAP-family polypeptide is modulated by administering to an individual, a therapeutically effective amount of a THAP-type chemokine-binding agent.

445. The method of Paragraph 444, wherein said THAP-type chemokine-binding agent comprises a therapeutically effective amount of a polypeptide selected from the group consisting of a THAP1 polypeptide, an chemokine-binding domain of a THAP1 polypeptide, a THAP1 polypeptide oligomer, an oligomer comprising a THAP1 chemokine-binding domain, a THAP1 polypeptide-immunoglobulin fusion, a THAP1 chemokine-binding domain-immunoglobulin fusion and polypeptide homologs having at least 30% amino acid identity to any one of the aforementioned polypeptides.

446. The method of Paragraph 445, wherein said chemokine-binding domain is an SLC-binding domain.

447. The method of Paragraph 445, wherein said chemokine-binding domain is a CXCL9-binding domain.

448. A method of reducing the symptoms associated with a condition resulting from the activity of a THAP-family polypeptide in an individual comprising modulating the extent of transcriptional repression or activation of at least one THAP-family responsive promoter in said individual.

449. The method of Paragraph 448, wherein said THAP-family polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114.

450. The method of Paragraph 448, wherein said THAP-family polypeptide comprises an amino acid sequence of SEQ ID NO: 3.

451. The method of Paragraph 448, wherein said THAP responsive promoter comprises a THAP responsive element.

452. The method of Paragraph 448, wherein said THAP responsive promoter does not comprise a THAP responsive element.

453. A method of reducing the symptoms associated with a condition resulting from the activity of a THAP-family polypeptide in an individual, said method comprising:

diagnosing said individual with a condition resulting from the activity of a THAP-family polypeptide; and

administering a compound which modulates the interaction between said THAP-family polypeptide and a chemokine to said individual.

454. The method of Paragraph 453, wherein said THAP-family polypeptide is selected from a group consisting of polypeptides having an amino acid sequence of SEQ ID NOs: 1-114.

455. The method of Paragraph 453, wherein said THAP-family polypeptide is THAP1.

456. The method of Paragraph 453, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

457. The method of Paragraph 453, wherein said chemokine is SLC.
458. The method of Paragraph 453, wherein said chemokine is CXCL9.
459. A method of reducing the symptoms associated with a condition resulting from the activity of a THAP-family polypeptide in an individual comprising:
- 5       diagnosing said individual with a condition resulting from the activity of THAP-family polypeptide; and
- administering a chemokine or an analog thereof to said individual.
460. The method of Paragraph 459, wherein said THAP-family polypeptide is selected from a group consisting of polypeptides having an amino acid sequence of SEQ ID NOs: 1-114.
- 10       461. The method of Paragraph 459, wherein said THAP-family polypeptide is THAP1.
462. The method of Paragraph 459, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.
463. The method of Paragraph 459, wherein said chemokine is SLC.
464. The method of Paragraph 459, wherein said chemokine is CXCL9.
- 15       465. A method of reducing the symptoms associated with transcriptional repression or activation mediated by a THAP-family polypeptide in an individual comprising administering a chemokine or an analog thereof to said individual.
466. The method of Paragraph 465, wherein said THAP-family polypeptide is selected from a group consisting of polypeptides having an amino acid sequence of SEQ ID NOs: 1-114.
- 20       467. The method of Paragraph 465, wherein said THAP-family polypeptide is THAP1.
468. The method of Paragraph 465, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.
469. The method of Paragraph 465, wherein said chemokine is SLC.
470. The method of Paragraph 465, wherein said chemokine is CXCL9.
- 25       471. A method of reducing the symptoms associated with the activity of a chemokine in an individual comprising modulating the extent to which said chemokine is transported to the nucleus of a cell in said individual.
472. The method of Paragraph 471, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.
- 30       473. The method of Paragraph 471, wherein said cell expresses a chemokine receptor selected from the group consisting of CCR1, CCR3, CCR5, CCR7, CCR11 and CXCR3.
474. The method of Paragraph 473, wherein said chemokine is SLC and said chemokine receptor is CCR7.
475. The method of Paragraph 473, wherein said chemokine is CXCL9 and said
- 35       chemokine receptor is CXCR3.

4/b. The method of Paragraph 471, wherein the extent of transport of said chemokine into a nucleus of a cell is modulated by contacting said chemokine with a THAP-type chemokine-binding agent.

477. The method of Paragraph 476, wherein said THAP-type chemokine-binding agent  
5 selected from the group consisting of a THAP1 polypeptide, a chemokine-binding domain of a THAP1 polypeptide, a THAP1 polypeptide oligomer, an oligomer comprising a THAP1 chemokine-binding domain, a THAP1 polypeptide-immunoglobulin fusion, a THAP1 chemokine-binding domain-immunoglobulin fusion and polypeptide homologs having at least 30% amino acid identity to any one of the aforementioned polypeptides.

10 478. The method of Paragraph 477, wherein said chemokine-binding domain is an SLC-binding domain.

479. The method of Paragraph 477, wherein said chemokine-binding domain is a CXCL9-binding domain.

480. A method for identifying a compound which modulates the transport of a  
15 chemokine into the nucleus comprising comparing the extent of said chemokine transport into the nucleus of cells in the presence and absence of a test compound.

481. The method of Paragraph 480, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

482. The method of Paragraph 480, wherein said cell expresses a chemokine receptor  
20 selected from the group consisting of CCR1, CCR3, CCR5, CCR7, CCR11 and CXCR3.

483. The method of Paragraph 482, wherein said chemokine is SLC and said chemokine receptor is CCR7.

484. The method of Paragraph 482, wherein said chemokine is CXCL9 and said chemokine receptor is CXCR3.

25 485. The method of Paragraph 480, wherein the extent of transport of said chemokine into a nucleus of a cell is modulated by contacting said chemokine with a THAP-type chemokine-binding agent.

486. The method of Paragraph 485, wherein said THAP-type chemokine-binding agent  
30 is selected from the group consisting of a THAP1 polypeptide, a chemokine-binding domain of a THAP1 polypeptide, a THAP1 polypeptide oligomer, an oligomer comprising a THAP1 chemokine-binding domain, a THAP1 polypeptide-immunoglobulin fusion, a THAP1 chemokine-binding domain-immunoglobulin fusion and polypeptide homologs having at least 30% amino acid identity to any one of the aforementioned polypeptides.

487. The method of Paragraph 486, wherein said chemokine-binding domain is an SLC-  
35 binding domain.

488. The method of Paragraph 486, wherein said chemokine-binding domain is a CXCL9-binding domain.

489. The method of Paragraph 480, wherein transport of SLC into the nucleus is measured by immunostaining.

490. A vector comprising a THAP responsive promoter operably linked to a nucleic acid encoding a detectable product.

5 491. The vector of Paragraph 490, wherein said THAP responsive promoter comprises a THAP responsive element.

492. The vector of Paragraph 490, wherein said THAP responsive promoter does not comprise a THAP responsive element.

10 493. A genetically engineered cell comprising the vector of any one of Paragraphs 490-492.

494. An *in vitro* transcription reaction comprising a nucleic acid comprising a THAP responsive promoter, ribonucleotides and an RNA polymerase.

495. The *in vitro* transcription reaction of Paragraph 494, wherein said THAP responsive promoter comprises a THAP responsive element.

15 496. An isolated mutant THAP-family polypeptide that does not bind to a chemokine.

497. The isolated mutant THAP-family polypeptide of Paragraph 496, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

20 498. The isolated mutant THAP-family polypeptide of Paragraph 496, wherein said chemokine is SLC.

499. The isolated mutant THAP-family polypeptide of Paragraph 496, wherein said chemokine is CXCL9.

500. The isolated mutant THAP-family polypeptide of Paragraph 496, wherein said THAP-family polypeptide is THAP1.

25 501. The isolated mutant THAP-family polypeptide of Paragraph 500, wherein said polypeptide comprises an amino acid sequence of SEQ ID NO: 3.

502. The isolated mutant THAP-family polypeptide of Paragraph 501, wherein said amino acid sequence comprises at least one point mutation.

30 503. The methods of Paragraphs 291, 310, 328, 358, 363, 388, or 401 or the compositions of Paragraphs 371 or 382, wherein said THAP-family polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOS: 1-114.

504. The methods of Paragraphs 294, 342, 358 or 363 or the compositions of Paragraphs 349 or 353, wherein said THAP responsive element comprises a nucleic acid having a nucleotide sequence selected from the group consisting of SEQ ID NOS: 140-159 and 306.

35 505. The methods of Paragraphs 291, 310, or 328, wherein said gene comprises a nucleic acid having a nucleotide sequence selected from the group consisting of SEQ ID NOS: 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384,

386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 530, 532, 534 and portions thereof.

5           506. The methods of Paragraphs 291, 310 or 328, wherein said gene encodes a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NOs: 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453, 455, 457, 459, 461,  
10       463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 531, 533, 535 and portions thereof.

507. The methods of Paragraphs 310, 328, 363, 388 or 401 or the composition of Paragraph 371, wherein said chemokine has an amino acid sequence selected from the group consisting of SEQ ID NOs: 119, 271, 273, 275, 277, 289 and 323.

15           508. A method of ameliorating symptoms associated with inflammation, said method comprising modulating the expression of a THAP responsive gene or a gene responsive to a THAP/chemokine complex.

509. The method of Paragraph 508, wherein said gene expression is modulated by modulating the interaction between a nucleic acid and a THAP-family polypeptide or a biologically  
20       active fragment thereof, modulating the interaction between a nucleic acid and a THAP/chemokine complex or modulating the interaction between a chemokine and THAP-family polypeptide or a biologically active fragment thereof.

510. A method of ameliorating symptoms associated with a condition resulting from excessive or insufficient angiogenesis, said method comprising modulating the expression of a  
25       THAP responsive gene or a gene responsive to a THAP/chemokine complex.

511. The method of Paragraph 510, wherein said gene expression is modulated by modulating the interaction between a nucleic acid and a THAP-family polypeptide or a biologically  
30       active fragment thereof, modulating the interaction between a nucleic acid and a THAP/chemokine complex or modulating the interaction between a chemokine and THAP-family polypeptide or a biologically active fragment thereof.

512. A method of ameliorating the symptoms associated with a condition resulting from the proliferation of a cancer cell, said method comprising modulating the expression of a THAP responsive gene or a gene responsive to a THAP/chemokine complex.

513. The method of Paragraph 512, wherein said gene expression is modulated by  
35       modulating the interaction between a nucleic acid and a THAP-family polypeptide or a biologically active fragment thereof, modulating the interaction between a nucleic acid and a THAP/chemokine

complex or modulating the interaction between a chemokine and THAP-family polypeptide or a biologically active fragment thereof.

It will be appreciated that THAP compositions and methods of making and using have been described in other copending patent applications. These patent applications include, US Patent  
5 Application No. 10/317,832, entitled NOVEL DEATH ASSOCIATED PROTEINS AND THAP1 AND PAR4 PATHWAYS IN APOPTOSIS CONTROL, filed December 10, 2002 and US Patent Application No. 10/601,072, entitled CHEMOKINE-BINDING PROTEIN AND METHODS OF USE, filed June 19, 2003.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

10 Figure 1A illustrates an amino acid sequence alignment of human THAP1 (hTHAP1) (SEQ ID NO: 3) and mouse THAP1 (mTHAP1) (SEQ ID NO: 99) orthologous polypeptides. Identical amino acid residues are indicated with an asterisk.

Figure 1B depicts the primary structure of the human THAP1 polypeptide. Positions of the THAP domain, the proline-rich region (PRO) and the bipartite nuclear localization sequence (NLS)  
15 are indicated.

Figure 2 depicts the results of a Northern Blot analysis of THAP1 mRNA expression in 12 human tissues. Each lane contains 2  $\mu$ g of poly A<sup>+</sup> RNA isolated from the indicated human tissues. The blot was hybridized, under high-stringency conditions, with a <sup>32</sup>P-labeled THAP1 cDNA probe, and exposed at -70°C for 72 hours.

20 Figure 3A illustrates the interaction between THAP1 and PAR4 in a yeast two-hybrid system. In particular, THAP1 binds to wild-type Par4 (Par4) and the leucine zipper-containing Par4 death domain (Par4DD) (amino acids 250-342 of PAR4) but not a Par4 deletion mutant lacking the death domain (PAR4 $\Delta$ ) (amino acids 1-276 of PAR4). A (+) indicates binding whereas a (-) indicated lack of binding.

25 Figure 3B shows the binding of *in vitro* translated, <sup>35</sup>S-methionine-labeled THAP1 to a GST-Par4DD polypeptide fusion. Par4DD was expressed as a GST fusion protein then purified on an affinity matrix of glutathione sepharose. GST served as negative control. The input represents 1/10 of the material used in the binding assay.

30 Figure 4A illustrates the interaction between PAR4 and several THAP1 deletion mutants both *in vitro* and *in vivo*. Each THAP1 deletion mutant was tested for binding to either PAR or PAR4DD in a yeast two hybrid system (two hybrid bait), to PAR4DD in GST pull down assays (*in vitro*) and to myc-Par4DD in primary human endothelial cells (*in vivo*). A (+) indicates binding whereas a (-) indicated lack of binding.

35 Figure 4B shows the binding of several *in vitro* translated, <sup>35</sup>S-methionine-labeled THAP1 deletion mutants to a GST-Par4DD polypeptide fusion. Par4DD was expressed as a GST fusion protein then purified on an affinity matrix of glutathione sepharose. GST served as negative control. The input represents 1/10 of the material used in the binding assay.



Figure 5A depicts an amino acid sequence alignment of the Par4 binding domain of human THAP1 (SEQ ID NO: 117) and mouse THAP1 (SEQ ID NO: 116) orthologues with that of mouse ZIP kinase (SEQ ID NO: 115), another Par4 binding partner. An arginine-rich consensus Par4 binding site (SEQ ID NO: 15), derived from this alignment, is also indicated.

5 Figure 5B shows the primary structure of the THAP1 wild-type polypeptide and two THAP1 mutants (THAP1 $\Delta$ (QRCRR) and THAP1 RR/AA). THAP1 $\Delta$ (QRCRR) is a deletion mutant having a deletion of amino acids at positions 168-172 of THAP1 (SEQ ID NO: 3) whereas THAP RR/AA is a mutant having the two arginines located at amino acid positions 171 and 172 to THAP1 (SEQ ID NO: 3) replaced with alanines. Results obtained, in yeast two-hybrid system with Par4 and Par4DD baits (two hybrid bait), in GST pull down assays with GST-Par4DD (*in vitro*) and in the *in vivo* interaction test with myc-Par4DD in primary human endothelial cells (*in vivo*) are summarized.

Figure 6A is a graph which compares apoptosis levels in cells transfected with GFP-APSK1, GFP-Par4 or GFP-THAP1 expression vectors. Apoptosis was quantified by DAPI staining of apoptotic nuclei, 24 h after serum-withdrawal. Values are the means of three independent experiments.

Figure 6B is a graph which compares apoptosis levels in cells transfected with GFP-APSK1 or GFP-THAP1 expression vectors. Apoptosis was quantified by DAPI staining of apoptotic nuclei, 24 h after addition of TNF  $\alpha$ . Values are the means of three independent experiments.

20 Figure 7A shows the binding of *in vitro* translated <sup>35</sup>S-methionine labeled THAP1 (wt) or THAP1 $\Delta$ THAP ( $\Delta$ ) to a GST-Par4DD polypeptide fusion. Par4DD was expressed as a GST fusion protein then purified on an affinity matrix of glutathione sepharose. GST served as negative control. The input represents 1/10 of the material used in the binding assay.

Figure 7B is a graph which compares the proapoptotic activity of THAP1 with a THAP1 mutant having its THAP domain (amino acids 1-90 of SEQ ID NO: 3) deleted. The percentage of apoptotic cells in mouse 3T3 fibroblasts overexpressing GFP-APSK1 (control), GFP-THAP1 (THAP1) or GFP-THAP1 $\Delta$ THAP (THAP1 $\Delta$ THAP) was determined by counting apoptotic nuclei after DAPI staining. Values are the means of three independent experiments.

30 Figure 8 depicts the primary structure of twelve human THAP proteins. The THAP domain (colored grey) is located at the amino-terminus of each of the twelve human THAP proteins. The black box in THAP1, THAP2 and THAP3 indicates a nuclear localization sequence, rich in basic residues, that is conserved in the three proteins. The number of amino-acids in each THAP protein is indicated; (\*) indicates the protein is not full length.

35 Figure 9A depicts an amino acid sequence alignment of the THAP domain of human THAP1 (hTHAP1, SEQ ID NO: 123) with the DNA binding domain of drosophila melanogaster P-element transposase (dmTransposase, SEQ ID NO: 124). Identical residues are boxed in black and conserved residues in grey. A THAP domain consensus sequence (SEQ ID NO: 125) is also shown.

Figure 9B depicts an amino acid sequence alignment of the THAP domains of twelve members of the human THAP family (hTHAP1, SEQ ID NO: 126; hTHAP2, SEQ ID NO: 131; hTHAP3, SEQ ID NO: 127; hTHAP4, SEQ ID NO: 130; hTHAP5, SEQ ID NO: 128; hTHAP6, SEQ ID NO: 135; hTHAP7, SEQ ID NO: 133; hTHAP8, SEQ ID NO: 129; hTHAP9, SEQ ID NO: 134; hTHAP10, SEQ ID NO: 137; hTHAP11, SEQ ID NO: 136; hTHAP0, SEQ ID NO: 132) with the DNA binding domain of *Drosophila melanogaster* P-element transposase (dmTransposase, SEQ ID NO: 138). Residues conserved among at least seven of the thirteen sequences are boxed. Black boxes indicate identical residues whereas boxes shaded in grey show similar amino acids. Dashed lines represent gaps introduced to align sequences. A THAP domain consensus sequence (SEQ ID NO: 139) is also shown.

Figure 9C depicts an amino acid sequence alignment of 95 distinct THAP domain sequences, including hTHAP1 through hTHAP11 and hTHAP0 (SEQ ID NOs: 3-14, listed sequentially beginning from the top), with 83 THAP domains from other species (SEQ ID NOs: 16-98, listed sequentially beginning at the sequence denoted sTHAP1 and ending at the sequence denoted ceNP\_498747.1), which were identified by searching GenBank genomic and EST databases with the human THAP sequences. Residues conserved among at least 50% of the sequences are boxed. Black boxes indicate identical residues whereas boxes shaded in grey show similar amino acids. Dashed lines represent gaps introduced to align sequences. The species are indicated: *Homo sapiens* (h); *Sus scrofa* (s); *Bos taurus* (b); *Mus musculus* (m); *Rattus norvegicus* (r); *Gallus gallus* (g); *Xenopus laevis* (x); *Danio rerio* (z); *Oryzias latipes* (o); *Drosophila melanogaster* (dm); *Anopheles gambiae* (a); *Bombyx mori* (bm); *Caenorhabditis elegans* (ce). A consensus sequence (SEQ ID NO: 2) is also shown. Amino acids underlined in the consensus sequence are residues which are conserved in all 95 THAP sequences.

Figure 10A shows an amino acid sequence alignment of the human THAP1 (SEQ ID NO: 3), THAP2 (SEQ ID NO: 4) and THAP3 (SEQ ID NO: 5) protein sequences. Residues conserved among at least two of the three sequences are boxed. Black boxes indicate identical residues whereas boxes shaded in grey show similar amino acids. Dashed lines represent gaps introduced to align sequences. Regions corresponding to the THAP domain, the PAR4-binding domain, and the nuclear localization signal (NLS) are also indicated.

Figure 10B shows the primary structure of human THAP1, THAP2 and THAP3 and results of two-hybrid interactions between each THAP protein and Par4 or Par4 death domain (Par4DD) in the yeast two hybrid system.

Figure 10C shows the binding of *in vitro* translated, <sup>35</sup>S-methionine-labeled THAP2 and THAP3 to a GST-Par4DD polypeptide fusion. Par4DD was expressed as a GST fusion protein then purified on an affinity matrix of glutathione sepharose. GST served as negative control. The input represents 1/10 of the material used in the binding assay.

Figure 11A is a graph which compares apoptosis levels in cells transfected with GFP-

APSK1, GFP-THAP2 or GFP-THAP3 expression vectors. Apoptosis was quantified by DAPI staining of apoptotic nuclei, 24 h after serum-withdrawal. Values are the means of two independent representative experiments.

Figure 11B is a graph which compares apoptosis levels in cells transfected with GFP-APSK1, GFP-THAP2 or GFP-THAP3 expression vectors. Apoptosis was quantified by DAPI staining of apoptotic nuclei, 24 h after additional of TNF $\alpha$ . Values are the means of two independent representative experiments.

Figure 12 illustrates the results obtained by screening several different THAP1 mutants in a yeast two-hybrid system with SLC/CCL21 bait. The primary structure of each THAP1 deletion mutant that was tested is shown. The 70 carboxy-terminal residues of THAP1 (amino acids 143-213) are sufficient for binding to chemokine SLC/CCL21.

Figure 13 illustrates the interaction of THAP1 with wild type SLC/CCL21 and a SLC/CCL21 mutant deleted of the basic carboxy-terminal extension (SLC/CCL21 $\Delta$ COOH). The interaction was analyzed both in yeast two-hybrid system with THAP1 bait and *in vitro* using GST-pull down assays with GST-THAP1.

Figure 14 depicts micrographs of the primary human endothelial cells were transfected with the GFP-THAP0, 1, 2, 3, 6, 7, 8, 10, 11 (green fluorescence) expression constructs. To reveal the nuclear localization of the human THAP proteins, nuclei were counterstained with DAPI (blue). The bar equals 5  $\mu$ m.

Figure 15A is a threading-derived structural alignment between the THAP domain of human THAP1 (THAP1) (amino acids 1-81 of SEQ ID NO: 3) and the thyroid receptor  $\beta$  DNA binding domain (NLLB) (SEQ ID NO: 121). The color coding is identical to that described in Figure 15D.

Figure 15B shows a model of the three-dimensional structure of the THAP domain of human THAP1 based on its homology with the crystallographic structure of thyroid receptor  $\beta$ . The color coding is identical to that described in Figure 15D.

Figure 15C shows a model of the three-dimensional structure of the DNA-binding domain of *Drosophila* transposase (DmTRP) based on its homology with the crystallographic structure of the DNA-binding domain of the glucocorticoid receptor. The color coding is identical to that described in Figure 15D.

Figure 15D is a threading-derived structural alignment between the *Drosophila melanogaster* transposase DNA binding domain (DmTRP) (SEQ ID NO: 120) and the glucocorticoid receptor DNA binding domain (GLUA) (SEQ ID NO: 122). In accordance with the sequences and structures in Figures 15A - 15C, the color-coding is the following: brown indicates residues in  $\alpha$ -helices; indigo indicates residues in  $\beta$ -strands; red denotes the eight conserved Cys residues in NLLB and GLUA or for the three Cys residues common to THAP1 and DmTRP;

magenta indicates other Cys residues in THAP1 or DmTRP; cyan denotes the residues involved in the hydrophobic interactions networks colored in THAP1 or DmTRP.

Figure 16A illustrates the results obtained by screening several different THAP1 mutants in a yeast two-hybrid system with THAP1 bait. The primary structure of each THAP1 deletion mutant that was tested is shown. A (+) indicates binding whereas a (-) indicates no binding.

Figure 16B shows the binding of several *in vitro* translated, <sup>35</sup>S-methionine-labeled THAP1 deletion mutants to a GST-THAP1 polypeptide fusion. Wild-type THAP1 was expressed as a GST fusion protein then purified on an affinity matrix of glutathione sepharose. GST served as negative control. The input represents 1/10 of the material used in the binding assay.

Figure 17A is an agarose gel showing two distinct THAP1 cDNA fragments were obtained by RT-PCR. Two distinct THAP1 cDNAs were ~400 and 600 nucleotides in length.

Figure 17B shows that the 400 nucleotide fragment corresponds to an alternatively spliced isoform of human THAP1 cDNA, lacking exon 2 (nucleotides 273-468 of SEQ ID 160).

Figure 17C is a Western blot which shows that the second isoform of human THAP1 (THAP1b) encodes a truncated THAP1 protein (THAP1 C3) lacking the amino-terminal THAP domain.

Figure 18A shows a specific DNA binding site recognized by the THAP domain of human THAP1. The THAP domain recognizes GGGCAA or TGGCAA DNA target sequences preferentially organized as direct repeats with 5 nucleotide spacing (DR-5). The consensus sequence 5'- GGGCAAnnnnnTGGCAA -3' (SEQ ID NO: 149). The DR-5 consensus was generated by examination of 9 nucleic acids bound by THAP1 (SEQ ID NO: 140-148, beginning sequentially from the top).

Figure 18B shows a second specific DNA binding site recognized by the THAP domain of human THAP1. The THAP domain recognizes everted repeats with 11 nucleotide spacing (ER-11) having a consensus sequence 5'- TTGCCAnnnnnnnnnnnGGGCAA -3' (SEQ ID NO: 159). The ER-11 consensus was generated by examination of 9 nucleic acids bound by THAP1 (SEQ ID NO: 150-158, beginning sequentially from the top).

Figure 19 shows that THAP1 interacts with both CC and CXC chemokines both *in vivo* in a yeast two-hybrid system with THAP1 prey and *in vitro* using GST-pull down assays with immobilized GST-THAP1. The cytokine IFN $\gamma$  was used as a negative control. Results are summarized as follows: +++ indicates strong binding; ++ indicates intermediate binding; +/- indicates some binding; - indicates no binding; and ND indicates not determined.

Figure 20A is an SDS-polyacrylamide gel showing the relative amounts of chemokine and cytokine used in immobilized GST-THAP1 binding assays.

Figure 20B is an SDS-polyacrylamide gel showing that neither the cytokine, IFN $\gamma$ , nor any of the chemokines bound to immobilized GST alone.

Figure 20C is an SDS-polyacrylamide gel showing that chemokines, CXCL10, CXCL9 and CCL19, but not the cytokine IFN $\gamma$ , bound to immobilized GST-THAP1 fusions.

Figure 21A shows the THAP1 protein fused to the Gal4 DNA-binding domain. This fusion was used in transcriptional assays with a *Gal*-UAS-luciferase reporter plasmid.

5 Figure 21B shows results of assays wherein the *Gal*-UAS-luciferase reporter plasmid was co-transfected into COS7 cells with increasing amounts of the Gal4 DNA-binding domain-THAP1 fusion expression vector. This analysis revealed that, compared to the Gal4 DNA-binding domain alone, the Gal4 DNA-binding domain-THAP1 fusion represses transcriptional activity of the luciferase reporter. The repression effect of THAP1 was similar to that observed with the well  
10 characterized transcriptional repressor Suv39H1.

Figure 22A shows that THAP1 as a nuclear receptor for chemokine SLC/CCL21. SLC binds to a cytoplasmic receptor such as CR7. Once internalized SLC/CCL21 is transported to the nucleus wherein it interacts with a THAP-family protein, such as THAP1. The bound SLC complex can bind DNA at certain recognition sequences so as to modulate transcription.

15 Figure 22B shows the role of THAP1 as a nuclear receptor for chemokines SLC/CCL21 and MIG/CXCL9. SLC and MIG bind to cell surface receptors such as CCR7 (polypeptide sequence SEQ ID NO: 302, nucleotide sequence SEQ ID NO: 303) and CXCR3 (polypeptide sequence SEQ ID NO: 304, nucleotide sequence SEQ ID NO: 305). Once internalized SLC and MIG are transported into the nucleus wherein they interact with a THAP-family protein, such as  
20 THAP1. The bound SLC/THAP1 and MIG/THAP1 complexes can bind DNA at certain recognition sequences so as to modulate transcription.

Figure 23 shows the nucleotide sequence of the human *Fucosyltransferase TVII* promoter (GenBank Accession Number AB012668, nucleotides 661-1080) (SEQ ID NO: 301). The sequence corresponding to the mRNA is underlined and the initiation codon (ATG) is indicated in  
25 bold. The promoter contains one GGGCAA (antisense orientation) and six GGGCAG (3 sense and 3 antisense orientations) THAP domain recognition elements, that are indicated in bold and underlined.

Figure 24 shows a consensus sequence (THAP-responsive element, THRE) (SEQ ID NO: 306) recognized by the THAP domain of human THAP1. The THRE consensus was generated by  
30 examination of 18 nucleic acids bound by THAP1 (SEQ ID NO: 140-148 and 150-158). The THRE was validated experimentally by using oligonucleotides mutated at each position.

Figure 25A shows the results of an EMSA assay carried out with the purified THAP domain from human THAP1 and oligonucleotides bearing wild type or mutant THRE sequences (wt, AGTAAGGGCAA (SEQ ID NO: 307); 3mut1, AGTAATTTCAG (SEQ ID NO: 308); 3mut3,  
35 AGTAAGGTCAA (SEQ ID NO: 309); 3mut4, AGTAAGTGCAA (SEQ ID NO: 310); 3mut14, AGTAAGGGCCA (SEQ ID NO: 311); and 3mut5, AGTAAGGGGAAA (SEQ ID NO: 312)).

Figure 25B shows the results of an EMSA assay carried out with the purified THAP domain from human THAP1 and labelled oligonucleotides bearing the wild type THRE sequence (5'-AGCAAGTAAGGGCAA**AACTACTTCAT**-3') (SEQ ID NO: 313) in the presence of increasing amounts of unlabelled THRE or non-specific competitor oligonucleotides (wild-type THRE, 5'-AGCAAGTAAGGGCAA**AACTACTTCAT**-3') (SEQ ID NO: 313) non-specific competitor, 5'-AGCAAGTAATTTCAA**AACTACTTCAT**-3') (SEQ ID NO: 314).

Figure 26A shows the results of an EMSA assay carried out with the purified THAP domain from human THAP1 and labelled oligonucleotides bearing the wild type THRE sequence (5'-AGCAAGTAAGGGCAA**AACTACTTCAT**-3') (SEQ ID NO: 313) in the presence of metal chelators EDTA (5mM or 50mM) or 1,10 phenanthroline (vehicle alone, 1mM or 5mM).

Figure 26B shows the results of an EMSA assay carried out with the purified THAP domain from human THAP1 and labelled oligonucleotides bearing the wild type THRE sequence (5'-AGCAAGTAAGGGCAA**AACTACTTCAT**-3') (SEQ ID NO: 313) in the presence of metal chelator 1,10 phenanthroline (5mM + Phe) and increasing amounts of Zn<sup>2+</sup> (100  $\mu$ M or 500  $\mu$ M) or Mg<sup>2+</sup> (100  $\mu$ M or 500  $\mu$ M).

Figure 27A-D depicts micrographs of human Hela cells transfected with the GFP-SLC (A) and GFP-MIG (green fluorescence) (C) expression constructs. To reveal the nuclear localization of the chemokines SLC and MIG, nuclei were counterstained with DAPI (blue) (B and D).

Figure 28A-D depicts micrographs of human U2OS cells transfected with the secreted MIG (red fluorescence) expression construct (phMIG-Flag) in the presence of a CXCR3 expression vector (pEF-CXCR3) (28C) or a control vector (pEF-puro) (28A). To reveal the nuclear localization of chemokine MIG, nuclei were counterstained with DAPI (blue) (B and D).

Figure 29A-C depicts micrographs of human U2OS cells transfected with the secreted MIG expression construct (phMIG-Flag) in the presence of a CXCR3 expression vector (pEF-CXCR3). MIG chemokine and CXCR3 expression were detected with anti-Flag (red fluorescence) (A) and anti-CXCR3 antibodies (green fluorescence) (B). To reveal the nuclear localization of chemokine MIG, nuclei were counterstained with DAPI (blue) (C).

Figure 30 shows the nucleotide sequence of the human *Survivin* promoter (GenBank Accession Number NT 010641.14, nucleotides 10102350-10102668) (SEQ ID NO: 315). The sequence corresponding to the mRNA is underlined and the initiation codon (ATG) is indicated in italics (nt 210-212). The promoter contains a DR5-type THAP1 responsive element in the antisense orientation (GGGCAAnnnnnGGGCAC) (SEQ ID NO: 316), that is indicated in bold.

Figure 31 shows the nucleotide sequence of the human *Ubiquitin specific protease 16* promoter (EPD database, which can be accessed by typing in the address bar of a web browser "http://www.epd." immediately followed by "isb-sib.ch"), Accession Number EP73421, nucleotides -499 to + 100) (SEQ ID NO: 317). The sequence corresponding to the mRNA is underlined. The promoter contains, near the TATA box, a consensus THAP1 responsive element

(THRE-11nt) in the antisense orientation (AGTGTGGGCAT) (SEQ ID NO: 318), that is indicated in bold and underlined.

### **DETAILED DESCRIPTION OF THE INVENTION**

#### *THAP and PAR4 biological pathways*

5 As mentioned above, the inventors have discovered a novel class of proteins involved in apoptosis. Then, the inventors have also linked a member of this novel class to another (PAR4) apoptosis pathway, and further linked both of these pathways to PML-NBs. Moreover, the inventors have also linked both of these pathways to endothelial cells, providing a range of novel and potentially selective therapeutic treatments. In particular, it has been discovered that THAP1  
10 (THanatos (death)-Associated-Protein-1) localizes to PML-NBs. Furthermore, two hybrid screening of an HEVEC cDNA library with the THAP1 bait lead to the identification of a unique interacting partner, the pro-apoptotic protein PAR4. PAR4 is also found to accumulate into PML-NBs. Targeting of the THAP-1 / PAR4 complex to PML-NBs is mediated by PML. Similarly to PAR4, THAP1 has a pro-apoptotic activity. This activity includes a novel motif in the amino-  
15 terminal part called THAP domain. Together these results define a novel PML-NBs pathway for apoptosis that involves the THAP1/PAR4 pro-apoptotic complex.

#### *THAP-family members, and uses thereof*

The present invention includes polynucleotides encoding a family of pro-apoptotic polypeptides THAP-0 to THAP11, and uses thereof for the modulation of apoptosis-related and  
20 other THAP-mediated activities. Included is THAP1, which forms a complex with the pro-apoptotic protein PAR4 and localizes in discrete subnuclear domains known as PML nuclear bodies. Additionally, THAP-family polypeptides can be used to alter or otherwise modulate bioavailability of SLC/CCL21 (SLC).

The present invention also includes a novel protein motif, the THAP domain, which is  
25 found in an 89 amino acid domain in the amino-terminal part of THAP1 and which is involved in THAP1 pro-apoptotic activity. The THAP domain defines a novel family of proteins, the THAP-family, with at least twelve distinct members in the human genome (THAP-0 to THAP11), which all contain a THAP domain in their amino-terminal part. The present invention thus pertains to nucleic acid molecules, including genomic and in particular the complete cDNA sequences,  
30 encoding members of the THAP-family, as well as with the corresponding translation products, nucleic acids encoding THAP domains, homologues thereof, nucleic acids encoding at least 10, 12, 15, 20, 25, 30, 40, 50, 100, 150 or 200 consecutive amino acids, to the extent that said span is consistent with the particular SEQ ID NO, of a sequence selected from the group consisting of SEQ ID NOs: 160-175.

35 THAP1 has been identified based on its expression in HEVs, specialized postcapillary venules found in lymphoid tissues and nonlymphoid tissues during chronic inflammatory diseases that support a high level of lymphocyte extravasation from the blood. An important element in the

cloning of the THAP1 cDNA from HEVECs was the development of protocols for obtaining HEVECs RNA, since HEVECs are not capable of maintaining their phenotype outside of their native environment for more than a few hours. A protocol was developed where total RNA was obtained from HEVECs freshly purified from human tonsils. Highly purified HEVECs were obtained by a combination of mechanical and enzymatic procedures, immunomagnetic depletion and positive selection. Tonsils were minced finely with scissors on a steel screen, digested with collagenase/dispase enzyme mix and unwanted contaminating cells were then depleted using immunomagnetic depletion. HEVECs were then selected by immunomagnetic positive selection with magnetic beads conjugated to the HEV-specific antibody MECA-79. From these HEVEC that were 98% MECA-79-positive, 1 µg of total RNA was used to generate full length cDNAs for THAP1 cDNA cloning and RT-PCR analysis.

As used herein, the term "nucleic acids" and "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. Throughout the present specification, the expression "nucleotide sequence" may be employed to designate indifferently a polynucleotide or a nucleic acid. More precisely, the expression "nucleotide sequence" encompasses the nucleic material itself and is thus not restricted to the sequence information (i.e. the succession of letters chosen among the four base letters) that biochemically characterizes a specific DNA or RNA molecule. Also, used interchangeably herein are terms "nucleic acids", "oligonucleotides", and "polynucleotides".

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated THAP-family nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having the nucleotide sequence of SEQ ID NOs: 160-175, a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NOs: 160-175, as a hybridization probe, THAP-family nucleic acid molecules can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. Molecular



Cloning. A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989).

Moreover, a nucleic acid molecule encompassing all or a portion of e.g. SEQ ID NOs: 160-175, can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NOs: 160-175.

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to THAP-family nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

As used herein, the term "hybridizes to" is intended to describe conditions for moderate stringency or high stringency hybridization, preferably where the hybridization and washing conditions permit nucleotide sequences at least 60% homologous to each other to remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85%, 90%, 95% or 98% homologous to each other typically remain hybridized to each other. Stringent conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are as follows: the hybridization step is realized at 65°C in the presence of 6 x SSC buffer, 5 x Denhardt's solution, 0.5% SDS and 100µg/ml of salmon sperm DNA. The hybridization step is followed by four washing steps:

- two washings during 5 min, preferably at 65°C in a 2 x SSC and 0.1%SDS buffer;
- one washing during 30 min, preferably at 65°C in a 2 x SSC and 0.1% SDS buffer,
- one washing during 10 min, preferably at 65°C in a 0.1 x SSC and 0.1%SDS buffer,

these hybridization conditions being suitable for a nucleic acid molecule of about 20 nucleotides in length. It will be appreciated that the hybridization conditions described above are to be adapted according to the length of the desired nucleic acid, following techniques well known to the one skilled in the art, for example be adapted according to the teachings disclosed in Hames B.D. and Higgins S.J. (1985) *Nucleic Acid Hybridization: A Practical Approach*. Hames and Higgins Ed., IRL Press, Oxford; and Current Protocols in Molecular Biolog (supra). Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence of SEQ ID NOs: 160-175 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the

sequence or a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, 90% or 95% of the length of the reference sequence (e.g., when aligning a second sequence to e.g. a THAP-1 amino acid sequence of SEQ ID NO: 3 having 213 amino acid residues, at least 50, preferably at least 100, more preferably at least 200, amino acid residues are aligned or when aligning a second sequence to the THAP-1 cDNA sequence of SEQ ID NO: 160 having 2173 nucleotides or nucleotides 202-835 which encode the amino acids of the THAP1 protein, preferably at least 100, preferably at least 200, more preferably at least 300, even more preferably at least 400, and even more preferably at least 500, 600, at least 700, at least 800, at least 900, at least 1000, at least 1200, at least 1400, at least 1600, at least 1800, or at least 2000 nucleotides are aligned. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % homology = number (#) of identical positions/total number (#) of positions 100).

The comparison of sequences and determination of percent homology between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Karlin and Altschul (1990) Proc. Natl. Acad. Sci. USA 87:2264-68, modified as in Karlin and Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-77. Such an algorithm is incorporated into the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) J. Mol. Biol. 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score=100, wordlength=12 to obtain nucleotide sequences homologous to THAP-family nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score=50, wordlength=3 to obtain amino acid sequences homologous to THAP-family protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) Nucleic Acids Research 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used (see, [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, CABIOS (1989). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN

program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

The term "polypeptide" refers to a polymer of amino acids without regard to the length of the polymer; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not specify or exclude post-expression modifications of polypeptides, for example, polypeptides which include the covalent attachment of glycosyl groups, acetyl groups, phosphate groups, lipid groups and the like are expressly encompassed by the term polypeptide. Also included within the definition are polypeptides which contain one or more analogs of an amino acid (including, for example, non-naturally occurring amino acids, amino acids which only occur naturally in an unrelated biological system, modified amino acids from mammalian systems etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of a protein according to the invention (e.g. THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof) in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of a protein according to the invention having less than about 30% (by dry weight) of protein other than the THAP-family protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of protein other than the protein according to the invention, still more preferably less than about 10% of protein other than the protein according to the invention, and most preferably less than about 5% of protein other than the protein according to the invention. When the protein according to the invention or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of a THAP-family protein having less than about 30% (by dry weight) of chemical precursors or non-THAP-family chemicals, more preferably less than about 20% chemical precursors or non-THAP-family or THAP-domain

chemicals, still more preferably less than about 10% chemical precursors or non-THAP-family or THAP-domain chemicals, and most preferably less than about 5% chemical precursors or non-THAP-family or THAP-domain chemicals.

5 The term "recombinant polypeptide" is used herein to refer to polypeptides that have been artificially designed and which comprise at least two polypeptide sequences that are not found as contiguous polypeptide sequences in their initial natural environment, or to refer to polypeptides which have been expressed from a recombinant polynucleotide.

Accordingly, another aspect of the invention pertains to anti-THAP-family or THAP-domain antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')<sub>2</sub> fragments which can be generated by treating the antibody with an enzyme such as pepsin. 15 The invention provides polyclonal and monoclonal antibodies that bind a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of a THAP-family or THAP domain polypeptide. A monoclonal antibody composition thus typically displays a single binding affinity for a particular THAP-family or THAP domain protein with which it immunoreacts. 20

#### PAR4

As mentioned above, Prostate apoptosis response-4 (PAR4) is a 38 kDa protein initially identified as the product of a gene specifically upregulated in prostate tumor cells undergoing apoptosis (for reviews see Rangnekar, 1998 ; Mattson et al., 1999). The PAR4 nucleic acid and amino acid sequences, see Johnstone et al, Mol. Cell. Biol. 16 (12), 6945-6956 (1996); and Genbank accession no. U63809 (SEQ ID NO: 118). 25

As used interchangeably herein, a "PAR4 activity", "biological activity of a PAR4" or "functional activity of a PAR4", refers to an activity exerted by a PAR4 protein, polypeptide or nucleic acid molecule as determined in vivo, or in vitro, according to standard techniques. In one embodiment, a PAR4 activity is a direct activity, such as an association with a PAR4-target molecule or most preferably apoptosis induction activity, or inhibition of cell proliferation or cell cycle. As used herein, a "target molecule" is a molecule with which a PAR4 protein binds or interacts in nature, such that PAR4-mediated function is achieved. An example of a PAR4 target molecule is a THAP-family protein such as THAP1 or THAP2, or a PML-NBs protein. A PAR4 target molecule can be a PAR4 protein or polypeptide or a non-PAR4 molecule. For example, a 35 PAR4 target molecule can be a non-PAR4 protein molecule. Alternatively, a PAR4 activity is an

indirect activity, such as an activity mediated by interaction of the PAR4 protein with a PAR4 target molecule such that the target molecule modulates a downstream cellular activity (e.g., interaction of a PAR4 molecule with a PAR4 target molecule can modulate the activity of that target molecule on an intracellular signaling pathway).

- 5 Binding or interaction with a PAR4 target molecule (such as THAP1/PAR4 described herein) or with other targets can be detected for example using a two hybrid-based assay in yeast to find drugs that disrupt interaction of the PAR4 bait with the target (e.g. PAR4) prey, or an in vitro interaction assay with recombinant PAR4 and target proteins (e.g. THAP1 and PAR4).

### CHEMOKINES

- 10 Chemokines are important in medicine because they regulate the movement and biological activities of leukocytes in many disease situations, including, but not limited to: allergic disorders, autoimmune diseases, ischemia/reperfusion injury, development of atherosclerotic plaques, cancer (including mobilization of hematopoietic stem cells for use in chemotherapy or myeloprotection during chemotherapy), chronic inflammatory disorders, chronic rejection of transplanted organs or  
15 tissue grafts, chronic myelogenous leukemia, and infection by HIV and other pathogens. Antagonists of chemokines or chemokine receptors may be of benefit in many of these diseases by reducing excessive inflammation and immune system responses.

- The activity of chemokines is tightly regulated to prevent excessive inflammation that can cause disease. Inhibition of chemokines by neutralizing antibodies in animal models (Sekido et al.  
20 (1993) *Nature* 365:654-657) or disruption of mouse chemokine genes (Cook et al. (1995) *Science* 269:1583-1588) have confirmed a critical role of chemokines in vivo in inflammation mediated by virus infection or other processes. The production of soluble versions of cytokine receptors containing only the extracellular binding domain, represents a physiological and therapeutic strategy to block the activity of some cytokines (Rose-John and Heinrich (1994) *Biochem J.*  
25 300:281-290; Heaney and Golde (1996) *Blood* 87:847-857). However, the seven transmembrane domain structure of chemokine receptors makes the construction of soluble, inhibitory receptors difficult, and thus antagonists based on mutated chemokines, blocking peptides or antibodies are under evaluation as chemokine inhibitors (D'Souza & Harden (1996) *Nature Medecine* 2:1293-1300; Howard et al. (1996) *Trends Biotech.* 14:46-51; Baggiolini (1998) *Nature* 392:565-568;  
30 Rollins (1997) *Blood* 90:909-928).

- Several viral chemokine binding proteins have been described that may be useful as soluble chemokine inhibitors. Soluble chemokine-binding proteins have been previously detected in poxviruses. Firstly, the myxoma virus T7 protein, which was first identified as a soluble IFN- $\gamma$  Receptor (Upton et al. (1992) *Science* 258:1369-1372), binds to a range of chemokines through the  
35 heparin-binding domain and affects the infiltration of cells into infected tissue (Lalani et al. (1997) *J Virol* 71:4356-4363). The protein is described in U.S. Patent No. 5,834,419 and International Publication No. WO 96/33730, and is designated CBP-1. Secondly, it was demonstrated that VV

strain *Listeria* expresses a soluble 35 kDa protein that is secreted from infected cells and which binds many CC chemokines. (Graham et al. (1997) *Virology* 229:12-24; Smith et al. (1997) *Virology* 236:316-327; Alcamí et al (1998) *J Immunol* 160:624-633), but not CXC chemokines, through a domain distinct from the heparin-binding domain (Smith et al. (1997) *Virology* 236:316-327; Alcamí et al (1998) *J Immunol* 160:624-633). This protein has been called vCKBP (Alcamí et al (1998) *J Immunol* 160:624-633). The protein is also described in U.S. Patent No. 5,871,740 and International Publication No. WO97/11714. One main disadvantage to the use of these viral proteins in a clinical setting is that antigenicity severely limits their indications. As such, there is a strong interest in the identification of cellular chemokine-binding proteins

Some aspects of the present invention relate to cellular polypeptides and homologs thereof, portions of cellular polypeptides and homologs thereof as well as modified cellular polypeptides and homologs thereof that bind to one or more chemokines. In some embodiments of the present invention such cellular polypeptides are THAP-family polypeptides, including THAP-1, chemokine-binding domains of THAP-family polypeptides (including a chemokine-binding domain of THAP-1), THAP-family polypeptide or THAP-family chemokine-binding domain fusions to immunoglobulin Fc (including THAP-1 fused to an immunoglobulin Fc region or a chemokine-binding domain of THAP-1 fused to an immunoglobulin Fc region), oligomers of THAP-family polypeptides or THAP-family chemokine-binding domains (including THAP-1 oligomers or oligomers of a chemokine-binding domain of THAP-1), or homologs of any of the above-listed compositions. Throughout this disclosure, the above-listed polypeptides are referred to as THAP-type chemokine-binding agents. Each of these THAP-type chemokine-binding agents are described in detail below.

### SLC/CCL21 (SLC)

#### *Biological Roles of SLC*

The signals which mediate T-cell infiltration during T-cell auto-immune diseases are poorly understood. SLC/CCL21 (SEQ ID NO: 119) is highly potent and highly specific for attracting T-cell migration. It was initially thought to be expressed only in secondary lymphoid organs, directing naive T-cells to areas of antigen presentation. However, using immunohistology it was found that expression of CCL21 was highly induced in endothelial cells of T-cell auto-immune infiltrative skin diseases (Christopherson et al. (2002) *Blood* electronic publication prior to printed publication). No other T-cell chemokine was consistently induced in these T cell skin diseases. The receptor for CCL21, CCR7, was also found to be highly expressed on the infiltrating T-cells, the majority of which expressed the memory CD45Ro phenotype. Inflamed venules endothelial cells expressing SLC/CCL21 in T cell infiltrative autoimmune skin diseases may therefore play a key role in the regulation of T-cell migration into these tissues.

There are a number of other autoimmune diseases where induced expression of SLC/CCL21 in endothelial cells may cause abnormal recruitment of T-cells from the circulation to

sites of pathologic inflammation. For instance, chemokine SLC/CCL21 appears to be important for aberrant T-cell infiltration in experimental autoimmune encephalomyelitis (EAE), an animal model for multiple sclerosis (Alt et al. (2002) *Eur J Immunol* 32:2133-44). Migration of autoaggressive T cells across the blood-brain barrier (BBB) is critically involved in the initiation of EAE. The direct involvement of chemokines in this process was suggested by the observation that G-protein-mediated signaling is required to promote adhesion strengthening of encephalitogenic T cells on BBB endothelium in vivo. A search for chemokines present at the BBB, by in situ hybridizations and immunohistochemistry revealed expression of the lymphoid chemokines CCL19/ELC and CCL21/SLC in venules surrounded by inflammatory cells (Alt et al. (2002) *Eur J Immunol* 32:2133-44). Their expression was paralleled by the presence of their common receptor CCR7 in inflammatory cells in brain and spinal cord sections of mice afflicted with EAE. Encephalitogenic T cells showed surface expression of CCR7 and specifically chemotaxed towards both CCL19 or CCL21 in a concentration dependent and pertussis toxin-sensitive manner comparable to naive lymphocytes in vitro. Binding assays on frozen sections of EAE brains demonstrated a functional involvement of CCL19 and CCL21 in adhesion strengthening of encephalitogenic T lymphocytes to inflamed venules in the brain (Alt et al. (2002) *Eur J Immunol* 32:2133-44). Taken together these data suggested that the lymphoid chemokines CCL19 and CCL21 besides regulating lymphocyte homing to secondary lymphoid tissue are involved in T lymphocyte migration into the immunoprivileged central nervous system during immunosurveillance and chronic inflammation.

Other diseases where induced expression of SLC/CCL21 in venular endothelial cells has been observed include rheumatoid arthritis (Page et al. (2002) *J Immunol* 168:5333-5341) and experimental autoimmune diabetes (Hjelmstrom et al. (2000) *Am J Pathol* 156:1133-1138). Therefore, chemokine SLC/CCL21 may be an important pharmacological target in T-cell autoimmune diseases. Inhibitors of SLC/CCL21 may be effective agents at treating these T cell infiltrative diseases by interfering with the abnormal recruitment of T cells, from the circulation to sites of pathologic inflammation, by endothelial cells expressing SLC/CCL21. The reduction in T cell migration into involved tissue would reduce the T-cell inflicted damage seen in those diseases.

Ectopic lymphoid tissue formation is a feature of many chronic inflammatory diseases, including rheumatoid arthritis, inflammatory bowel diseases (Crohn's disease, ulcerative colitis), autoimmune diabetes, chronic inflammatory skin diseases (lichen planus, psoriasis, ...), Hashimoto's thyroiditis, Sjogren's syndrome, gastric lymphomas and chronic inflammatory liver disease (Girard and Springer (1995) *Immunol today* 16:449-457; Takemura et al. (2001) *J Immunol* 167:1072-1080; Grant et al. (2002) *Am J Pathol* 2002 160:1445-55; Yoneyama et al. (2001) *J Exp Med* 193:35-49).

Ectopic expression of SLC/CCL21 has been shown to induce lymphoid neogenesis, both in mice and in human inflammatory diseases. In mice, transgenic expression of SLC/CCL21 in the pancreas (Fan et al. (2000) *J Immunol* 164:3955-3959; Chen et al. (2002) *J Immunol* 168:1001-

1008; Luther et al. (2002) J Immunol 169:424-433), a non-lymphoid tissue, has been found to be sufficient for the development and organization of ectopic lymphoid tissue through differential recruitment of T and B lymphocytes and induction of high endothelial venules, specialized blood vessels for lymphocyte migration (Girard and Springer (1995) Immunol today 16:449-457). In  
5 humans, hepatic expression of SLC/CCL21 has been shown to promote the development of high endothelial venules and portal-associated lymphoid tissue in chronic inflammatory liver disease (Grant et al. (2002) Am J Pathol 2002 160:1445-55; Yoneyama et al. (2001) J Exp Med 193:35-49). The chronic inflammatory liver disease primary sclerosing cholangitis (PSC) is associated with portal inflammation and the development of neolymphoid tissue in the liver. More than 70% of  
10 patients with PSC have a history of inflammatory bowel disease and strong induction of SLC/CCL21 on CD34(+) vascular endothelium in portal associated lymphoid tissue in PSC has been reported (Grant et al. (2002) Am J Pathol 2002 160:1445-55). In contrast, CCL21 is absent from LYVE-1(+) lymphatic vessel endothelium. Intrahepatic lymphocytes in PSC include a population of CCR7(+) T cells only half of which express CD45RA and which respond to CCL21  
15 in migration assays. The expression of CCL21 in association with mucosal addressin cell adhesion molecule-1 in portal tracts in PSC may promote the recruitment and retention of CCR7(+) mucosal lymphocytes leading to the establishment of chronic portal inflammation and the expanded portal-associated lymphoid tissue. These findings are supported by studies in an animal model of chronic hepatic inflammation, that have shown that anti-SLC/CCL21 antibodies prevent the development of  
20 high endothelial venules and portal-associated lymphoid tissue (Yoneyama et al. (2001) J Exp Med 193:35-49).

Induction of chemokine SLC/CCL21 at a site of inflammation could convert the lesion from an acute to a chronic state with corresponding development of ectopic lymphoid tissue. Blocking chemokine SLC/CCL21 activity in chronic inflammatory diseases may therefore have  
25 significant therapeutic value.

*Chemokine SLC/CCL21 and regulation of cell proliferation and cell death*

In addition to its key role in chemotaxis and cell migration, chemokine SLC/CCL21 has also been shown to regulate cell proliferation and cell death. For instance, the proliferation rate of normal hematopoietic or leukemia progenitor cells was reduced upon stimulation with SLC/CCL21  
30 (Hromas et al. (1997) J Immunol 159 :2554-2558 ; Hromas et al. (2000) Blood 95 :1506-1508). In contrast, SLC/CCL21 stimulated proliferation of mesangial cells from human kidney (Banas et al. (2002) J Immunol 168 :4301-4307), suggesting differential action of this chemokine on hematopoietic or non-hematopoietic cells.

SLC/CCL21 has also been shown to inhibit cell death. It was found that pretreatment with  
35 small doses of SLC/CCL21 prevented the death of normal murine marrow progenitors from the toxic effects of the chemotherapeutic agent Ara-C (Hromas et al. (2002) Cancer Chemother Pharmacol 50 :163-166). In addition, SLC/CCL21 was found to act as anti-apoptotic factor that



promotes mesengial cells survival in cell death assays. It is not known whether SLC/CCL21 effects on cell proliferation and cell death require the CCR7 chemokine receptor or are mediated by other cellular receptors.

5 *Chemokine SLC/CCL21 and regulation of endothelial cell differentiation (induction of the specialized high endothelial venule phenotype)*

Chemokine SLC/CCL21 has been shown to act on endothelial cells in two ways: 1) It exhibits angiostatic (anti-angiogenic) properties and efficiently block blood vessel formation *in vivo* (Soto et al. (1998) PNAS 95:8205-8210; Vicari et al. (2000) 165:1992-2000); 2) It induces differentiation of 'flat' endothelial cells into high endothelial venules (HEV), specialized blood vessels for lymphocyte migration (Girard and Springer (1995) Immunol today 16:449-457). For instance, in transgenic mice, ectopic expression of SLC/CCL21 in the pancreas (Fan et al. (2000) J Immunol 164:3955-3959; Chen et al. (2002) J Immunol 168:1001-1008; Luther et al. (2002) J Immunol 169:424-433), has been found to be sufficient for induction of high endothelial venules and associated lymphoid tissue. In humans, hepatic expression of SLC/CCL21 has been shown to promote the development of high endothelial venules and portal-associated lymphoid tissue in chronic inflammatory liver disease (Grant et al. (2002) Am J Pathol 2002 160:1445-55; Yoneyama et al. (2001) J Exp Med 193:35-49). A critical role for SLC/CCL21 in induction of high endothelial venules is supported by studies in an animal model of chronic hepatic inflammation, that have shown that anti-SLC/CCL21 antibodies prevent the development of high endothelial venules and portal-associated lymphoid tissue (Yoneyama et al. (2001) J Exp Med 193:35-49).

Induction of chemokine SLC/CCL21 at a site of inflammation might convert the lesion from an acute to a chronic state with corresponding development of high endothelial venules and ectopic lymphoid tissue. Blocking chemokine SLC/CCL21 effects on endothelial cells in chronic inflammatory diseases may therefore have significant therapeutic value. Since the CCR7 chemokine receptor is not expressed in endothelial cells, the effects of SLC/CCL21 on endothelial cells are likely to be mediated by other mechanisms. There is therefore a strong interest in the identification of other cellular receptors for SLC/CCL21.

**CHEMOKINES MIG/CXCL9, IP10/CXCL10, I-TAC/CXCL11**

*Roles of chemokines MIG/CXCL9, IP10/CXCL10, I-TAC/CXCL11 in leukocyte chemotaxis*

30 Chemokines monokine induced by IFN- $\gamma$  (Mig/CXCL9), IFN-induced protein of 10 kDa (IP-10/CXCL10) and IFN-inducible T cell  $\alpha$ -chemoattractant (I-TAC/CXCL11) are three CXC chemokines more closely related to each other than to any other chemokine with an amino acid sequence identity of about 40% (Luster and Ravetch (1987) J Exp Med 166:1084; Cole et al. (1998) J Exp Med 187 :2009-2021; Farber (1993) BBRC 192:223-230). They share a number of features:

35 i) they lack the glutamic acid - leucine - arginine (ELR) motif preceding the first conserved cysteine and are therefore inactive towards neutrophils; ii) they share an individual branch of the phylogenetic tree, have a similar gene structure, and are clustered on chromosome 4q21.2

(O'Donovan et al. (1999) Cytogenet Cell Genet 84:39-42). Among the CXC members, CXCL9, CXCL10 and CXCL11 are unique in that they are all induced by IFN- $\gamma$  in a wide variety of cell types, including endothelial cells (Luster and Ravetch (1987) J Exp Med 166:1084; Farber (1997) J Leuk Biol 61:246-257; Mach et al. (1999) J Clin Invest 104:1041; Cole et al. (1998) J Exp Med 187:2009-2021; Loetscher et al. (1998) Eur J Immunol 28:3696-3705), and act through a unique chemokine receptor, CXCR3. CXCR3 is expressed on activated T cells, preferentially of the Th1 phenotype, NK cells, and on a significant fraction (~20-40%) of circulating CD4<sup>+</sup> and CD8<sup>+</sup> T cells (Loetscher et al. (1996) J Exp Med 184:963-969; Loetscher et al. (1998) Eur J Immunol 28:3696-3705). The majority of peripheral CXCR3<sup>+</sup> T cells express CD45RO (memory T cells) as well as  $\beta_1$  integrins (Qin et al. (1998) J Clin Invest 101:746) which are implicated in the binding of lymphocytes to endothelial cells and the extracellular matrix. In addition, CXCR3 has been reported to be expressed on plasmacytoid dendritic cells, leukemic B cells, eosinophils, and dividing microvascular endothelial cells (Cella et al. (1999) Nat Med 5:919; Romagnani et al. (2001) J Clin Invest 107:53).

CXCR3<sup>+</sup> T cells accumulate at sites of Th1-type inflammation where IFN- $\gamma$  is highly expressed, including atherosclerosis, sarcoidosis, inflammatory bowel diseases, and rheumatoid arthritis (Qin et al. (1998) J Clin Invest 101:746; Mach et al. (1999) J Clin Invest 104:1041). IP-10 has been found to be highly expressed in a number of Th1-type inflammatory diseases, including psoriasis, tuberculoid leprosy, sarcoidosis, and viral meningitis. In addition, IFN- $\gamma$ -stimulated endothelial cells and endothelium from atherosclerotic lesions are a rich source of IP-10, Mig, and I-TAC suggesting an important role for these chemokines in the transendothelial migration and local retention of CXCR3<sup>+</sup> T cells found in atherosclerotic lesions (Mach et al. (1999) J Clin Invest 104:1041). In support of this hypothesis, IP-10 and Mig induce the rapid adhesion of IL-2-activated T cells to immobilized VCAM-1 and ICAM-1, and IP-10, Mig, and I-TAC are potent chemotactic agents for activated T cells.

*Roles of chemokines MIG/CXCL9, IP10/CXCL10, I-TAC/CXCL11 in angiogenesis*

CXC chemokines MIG/CXCL9, IP10/CXCL10, I-TAC/CXCL11 exhibit the selective property to inhibit angiogenesis (Belperio et al. (2000) J Leukoc Biol 68:1-8). These angiostatic chemokines induce injury to established tumor-associated vasculature and promote extensive tumor necrosis (Arenberg et al. (1996) J Exp Med 184:981-992; Sgadari et al. (1997) Blood 89:2635-2643) and thus have been proposed as useful therapeutic agents in cancer.

The angiostatic effects of CXCL9, CXCL10, and CXCL11 on human microvascular endothelial cells (HMVEC) are mediated by CXCR3 (Romagnani et al. (2001) J Clin Invest 107:53-63; Lasagni et al. (2003) J Exp Med 197:1537-1549). A distinct, previously unrecognized alternatively spliced variant of CXCR3 named CXCR3-B, has recently been shown to mediate the angiostatic activity of CXCR3 ligands (Lasagni et al. (2003) J Exp Med 197:1537-1549). Human microvascular endothelial cell line-1 (HMEC-1), transfected with either the known CXCR3

(renamed CXCR3-A) or CXCR3-B, bound CXCL9, CXCL10, and CXCL11. Overexpression of CXCR3-A induced an increase of survival, whereas overexpression of CXCR3-B dramatically reduced DNA synthesis and up-regulated apoptotic HMEC-1 death through activation of distinct signal transduction pathways. Unlike CXCR3A, CXCR3B was not found to be coupled to G-proteins. Remarkably, primary cultures of human microvascular endothelial cells, whose growth is inhibited by CXCL9, CXCL10 and CXCL11, expressed CXCR3-B, but not CXCR3-A. Finally, monoclonal antibodies raised to selectively recognize CXCR3-B reacted with endothelial cells from neoplastic tissues, providing evidence that CXCR3-B is also expressed *in vivo* and may account for the angiostatic effects of CXC chemokines.

10 *Chemokine MIG/CXCL9 and chemokine receptor CXCR3 and regulation of endothelial cell differentiation (induction of the specialized high endothelial venule phenotype)*

During inflammation, chemokine MIG/CXCL9 has been shown to be induced in high endothelial venules (HEV, Girard and Springer (1995) *Immunol today* 16:449-457), specialized blood vessels for lymphocyte migration (Janatpour et al. (2001) *J Exp Med* 193:1375-1384). Interestingly, in many human chronic inflammatory diseases, including Crohn's disease, Graves's disease and glomerulonephritis, CXCR3 receptor has also been found to be upregulated on endothelial cells during transformation of small blood vessels into HEV-like vessels (Romagnani et al. (2001) *J Clin Invest* 107:53-63).

20 Induction of chemokine MIG/CXCL9 and its receptor CXCR3 on endothelial cells at a site of inflammation might convert the lesion from an acute to a chronic state with corresponding development of high endothelial venules and ectopic lymphoid tissue. Blocking chemokine MIG/CXCL9 effects on CXCR3+ endothelial cells in chronic inflammatory diseases may therefore have significant therapeutic value.

*Role of chemokines CXCL9/Mig and CXCL10/IP-10 in vascular pericyte proliferation*

25 CXCL9 and CXCL10 have been implicated in the pathogenesis of proliferative glomerulonephritis, a common renal disease characterized by glomerular hypercellularity, because they induce increased survival and growth of human mesangial cells (HMC) through their receptor CXCR3 (Romagnani et al. (1999) *J Am Soc Nephrol* 10:2518-2526; Romagnani et al. (2002) *J Am Soc Nephrol* 13:53-64). High levels of expression of mRNA and protein for CXCL10 and CXCL9 were observed, by using *in situ* hybridization and immunohistochemical analyses, in kidney biopsy specimens from patients with glomerulonephritis (GN), particularly those with membranoproliferative or crescentic GN, but not in normal kidneys (Romagnani et al. (2002) *J Am Soc Nephrol* 13:53-64). Double-immunostaining or combined *in situ* hybridization and immunohistochemical analyses for IP-10, Mig, and proliferating cell nuclear antigen (PCNA) or  $\alpha$ -smooth muscle actin ( $\alpha$ -SMA) revealed that IP-10 and Mig production by resident glomerular cells was a selective property of glomeruli in which mesangial cells demonstrated active proliferation. IP-10 and Mig mRNA and protein were also expressed by primary cultures of human mesangial cells.

Moreover, high levels of CXCR3 were found in mesangial cells from patients with proliferative GN, and CXCR3 was also observed on the surface of cultured human mesangial cells (HMC) and seemed to mediate both intracellular  $\text{Ca}^{2+}$  influx, cell chemotaxis and cell proliferation, induced by CXCL9 and CXCL10 (Romagnani et al. (1999) J Am Soc Nephrol 10:2518-2526). Therefore, among patients with proliferative GN, the chemokines IP-10 and/or Mig not only may be responsible for the attraction of infiltrating mononuclear cells into the inflamed tissue but also may directly stimulate the proliferation of mesangial cells.

As used herein, "SLC/CCL21" and "SLC" are synonymous.

As used herein, "ELC/CCL19", "CCL19" and "ELC" are synonymous.

As used herein, "Rantes/CCL5", "CCL5" and "Rantes" are synonymous.

As used herein, "MIG/CXCL9", "CXCL9" and "MIG" are synonymous.

As used herein, "IP10/CXCL10", "CXCL10" and "IP10" are synonymous.

As used herein, "I-TAC/CXCL11", "CXCL11" and "I-TAC" are synonymous.

As used herein, in some embodiments of the present invention, "CXCR3" includes CXCR3 splice variant B (polypeptide encoding CXCR3 splice variant B, SEQ ID NO: 517; cDNA encoding CXCR3 splice variant B, Genbank Accession Number: AX805367, SEQ ID NO: 518).

#### **THAP-family members comprising a THAP Domain**

Based on the elucidation of a biological activity of the THAP1 protein in apoptosis as described herein, the inventors have identified and further characterized a novel protein motif, referred to herein as THAP domain. The THAP domain has been identified by the present inventors in several other polypeptides, as further described herein. Knowledge of the structure and function of the THAP domain allows the performing of screening assays that can be used in the preparation or screening of medicaments capable of modulating interaction with a THAP-family-target molecule, modulating cell cycle and cell proliferation, inducing apoptosis or enhancing or participating in the induction of apoptosis.

As used interchangeably herein, a THAP-family protein or polypeptide, or a THAP-family member refers to any polypeptide having a THAP domain as described herein. As mentioned, the inventors have provided several specific THAP-family members. Thus, as referred to herein, a THAP-family protein or polypeptide, or a THAP-family member, includes but is not limited to a THAP-0, THAP1, THAP-2, THAP-3, THAP-4, THAP-5, THAP-6, THAP-7, THAP-8, THAP-9, THAP10 or a THAP11 polypeptide.

As used interchangeably herein, a "THAP-family activity", "biological activity of a THAP-family member" or "functional activity of a THAP-family member", refers to an activity exerted by a THAP family or THAP domain polypeptide or nucleic acid molecule, or a biologically active fragment or homologue thereof comprising a THAP as determined in vivo, or in vitro, according to standard techniques. In one embodiment, a THAP-family activity is a direct activity, such as an association with a THAP-family-target molecule or most preferably apoptosis induction activity, or

inhibition of cell proliferation or cell cycle. As used herein, a "THAP-family target molecule" is a molecule with which a THAP-family protein binds or interacts in nature, such that a THAP family-mediated function is achieved. For example, a THAP family target molecule can be another THAP-family protein or polypeptide which is substantially identical or which shares structural similarity (e.g. forming a dimer or multimer). In another example, a THAP family target molecule can be a non-THAP family comprising protein molecule, or a non-self molecule such as for example a Death Domain receptor. Binding or interaction with a THAP family target molecule (such as THAP1/PAR4 described herein) or with other targets can be detected for example using a two hybrid-based assay in yeast to find drugs that disrupt interaction of the THAP family bait with the target (e.g. PAR4) prey, or an in vitro interaction assay with recombinant THAP family and target proteins (e.g. THAP1 and PAR4). In yet another example, a THAP family target molecule can be a nucleic acid molecule. For instance, a THAP family target molecule can be DNA.

Alternatively, a THAP-family activity may be an indirect activity, such as an activity mediated by interaction of the THAP-family protein with a THAP-family target molecule such that the target molecule modulates a downstream cellular activity (e.g., interaction of a THAP-family molecule with a THAP-family target molecule can modulate the activity of that target molecule on an intracellular signaling pathway).

THAP-family activity is not limited to the induction of apoptotic activity, but may also involve enhancing apoptotic activity. As death domains may mediate protein-protein interactions, including interactions with other death domains, THAP-family activity may involve transducing a cytotoxic signal.

Assays to detect apoptosis are well known. In a preferred example, an assay is based on serum-withdrawal induced apoptosis in a 3T3 cell line with tetracycline-regulated expression of a THAP family member comprising a THAP domain. Other non-limiting examples are also described.

In one example, a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof can be the minimum region of a polypeptide that is necessary and sufficient for the generation of cytotoxic death signals. Exemplary assays for apoptosis activity are further provided herein.

In specific embodiments, PAR4 is a preferred THAP1 and/or THAP2 target molecule. In another aspect, a THAP1 target molecule is a PML-NB protein.

In further aspects, THAP-domain or a THAP-family polypeptide comprises a DNA binding domain.

In other aspects, a THAP-family activity is detected by assessing any of the following activities: (1) mediating apoptosis or cell proliferation when expressed in or introduced into a cell, most preferably inducing or enhancing apoptosis, and/or most preferably reducing cell proliferation; (2) mediating apoptosis or cell proliferation of an endothelial cell; (3) mediating apoptosis or cell

proliferation of a hyperproliferative cell; (4) mediating apoptosis or cell proliferation of a CNS cell, preferably a neuronal or glial cell; (5) an activity determined in an animal selected from the group consisting of mediating, preferably inhibiting angiogenesis, mediating, preferably inhibiting inflammation, inhibition of metastatic potential of cancerous tissue, reduction of tumor burden, increase in sensitivity to chemotherapy or radiotherapy, killing a cancer cell, inhibition of the growth of a cancer cell, or induction of tumor regression; or (6) interaction with a THAP family target molecule or THAP domain target molecule, preferably interaction with a protein or a nucleic acid. Detecting THAP-family activity may also comprise detecting any suitable therapeutic endpoint discussed herein in the section titled "Methods of Treatment". THAP-family activity may be assessed either in vitro (cell or non-cell based) or in vivo depending on the assay type and format.

A THAP domain has been identified in the N-terminal region of the THAP1 protein, from about amino acid 1 to about amino acid 89 of SEQ ID NO: 3 based on sequence analysis and functional assays. A THAP domain has also been identified in THAP2 to THAP0 of SEQ ID NOs: 4-14. However, it will be appreciated that a functional THAP domain may be only a small portion of the protein, about 10 amino acids to about 15 amino acids, or from about 20 amino acids to about 25 amino acids, or from about 30 amino acids to about 35 amino acids, or from about 40 amino acids to about 45 amino acids, or from about 50 amino acids to about 55 amino acids, or from about 60 amino acids to about 70 amino acids, or from about 80 amino acids to about 90 amino acids, or about 100 amino acids in length. Alternatively, THAP domain or THAP family polypeptide activity, as defined above, may require a larger portion of the native protein than may be defined by protein-protein interaction, DNA binding, cell assays or by sequence alignment. A portion of a THAP domain-containing polypeptide from about 110 amino acids to about 115 amino acids, or from about 120 amino acids to 130 amino acids, or from about 140 amino acids to about 150 amino acids, or from about 160 amino acids to about 170 amino acids, or from about 180 amino acids to about 190 amino acids, or from about 200 amino acids to about 250 amino acids, or from about 300 amino acids to about 350 amino acids, or from about 400 amino acids to about 450 amino acids, or from about 500 amino acids to about 600 amino acids, to the extent that said length is consistent with the SEQ ID No, or the full length protein, for example any full length protein in SEQ ID NOs: 1-14, may be required for function.

As discussed, the invention includes a novel protein domain, including several examples of THAP-family members. The invention thus encompasses a THAP-family member comprising a polypeptide having at least a THAP domain sequence in the protein or corresponding nucleic acid molecule, preferably a THAP domain sequence corresponding to SEQ ID NOs: 1-2. A THAP-family member may comprise an amino acid sequence of at least about 25, 30, 35, 40, 45, 50, 60, 70, 80 to 90 amino acid residues in length, of which at least about 50-80%, preferably at least about

60-70%, more preferably at least about 65%, 75% or 90% of the amino acid residues are identical or similar amino acids to the THAP consensus domain SEQ ID NOs: 1-2.

Identity or similarity may be determined using any desired algorithm, including the algorithms and parameters for determining homology which are described herein.

5        Optionally, a THAP-domain-containing THAP-family polypeptide comprises a nuclear localization sequence (NLS). As used herein, the term nuclear localization sequence refers to an amino sequence allowing the THAP-family polypeptide to be localized or transported to the cell nucleus. A nuclear localization sequence generally comprises at least about 10, preferably about 13, preferably about 16, more preferably about 19, and even more preferably about 21, 23, 25, 30, 35 or  
10    40 amino acid residues. Alternatively, a THAP-family polypeptide may comprise a deletion of part or the entire NLS or a substitution or insertion in a NLS sequence, such that the modified THAP-family polypeptide is not localized or transported to the cell nucleus.

Isolated proteins of the present invention, preferably THAP family or THAP domain polypeptides, or a biologically active fragments or homologues thereof, have an amino acid  
15    sequence sufficiently homologous to the consensus amino acid sequence of SEQ ID NOs: 1-2. As used herein, the term "sufficiently homologous" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share  
20    common structural domains or motifs and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains have at least about 30-40% identity, preferably at least about 40-50% identity, more preferably at least about 50-60%, and even more preferably at least about 60-70%, 70-80%, 80%, 90%, 95%, 97%, 98%, 99% or 99.8% identity across the amino acid sequences of the domains and contain at least one and preferably two  
25    structural domains or motifs, are defined herein as sufficiently homologous. Furthermore, amino acid or nucleotide sequences which share at least about 30%, preferably at least about 40%, more preferably at least about 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99% or 99.8% identity and share a common functional activity are defined herein as sufficiently homologous.

It be appreciated that the invention encompasses any of the THAP-family polypeptides, as  
30    well as fragment thereof, nucleic acids complementary thereto and nucleic acids capable of hybridizing thereto under stringent conditions.

As used herein, "THAP/chemokine complex" refers to a THAP-family polypeptide or a biologically active fragment thereof in association with a chemokine or a biologically active fragment thereof. In some embodiments, THAP/chemokine complexes include, but are not limited to, THAP1/SLC, THAP1/MIG, THAP1/CXCL10, THAP1/CXCL11, THAP1/CCL19 and  
35    THAP1/CCL5.

#### **THAP-0 to THAP11**

As mentioned, the inventors have identified several THAP-family members, including THAP-0, THAP1, THAP-2, THAP-3, THAP-4, THAP-5, THAP-6, THAP-7, THAP-8, THAP-9, THAP10 and THAP11.

#### THAP1 Nucleic Acids

5 The human THAP1 coding sequence, which is approximately 639 nucleotides in length shown in SEQ ID NO: 160, encodes a protein which is approximately 213 amino acid residues in length. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP1 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic  
10 methods and drug screening assays as further described herein.

The human THAP1 gene is localized at chromosomes 8, 18, 11.

The THAP1 protein comprises a THAP domain at amino acids 1-89, the role of which in apoptosis is further demonstrated herein. The THAP1 protein comprises an interferon gamma homology motif at amino acids 136-169 of human THAP1  
15 (NYTVEDTMHQKRRIHQLEQQVEKLRKKLKTAAQR) (SEQ ID NO: 178), exhibiting 41% identity in a 34 residue overlap with human interferon gamma (amino acids 98-131). PML-NBs are closely linked to IFN gamma, and many PML-NB components are induced by IFN gamma, with IFN gamma responsive elements in the promoters of the corresponding genes. The THAP1 protein also includes a nuclear localization sequence at amino acids 146-165 of human THAP1  
20 (RKRIHQLEQQVEKLRKKLKT) (SEQ ID NO: 179). This sequence is responsible for localization of THAP1 in the nucleus. As demonstrated in the examples provided herein, deletion mutants of THAP1 lacking this sequence are no longer localized in the cell nucleus. The THAP1 protein further comprises a PAR4 binding motif (LE(X)<sub>14</sub> QRXRRQXR(X)<sub>11</sub> QR/KE) (SEQ ID NO: 180). The core of this motif has been defined experimentally by site directed mutagenesis and by  
25 comparison with mouse ZIP/DAP-like kinase (another PAR4 binding partner) it overlaps amino acids 168-175 of human THAP1 but the motif may also include a few residues upstream and downstream.

ESTs corresponding to THAP1 have been identified, and may be specifically included or excluded from the nucleic acids of the invention. The ESTs, as indicated below by accession  
30 number, provide evidence for tissue distribution for THAP1 as follows: AL582975 (B cells from Burkitt lymphoma); BG708372 (Hypothalamus); BG563619 (liver); BG497522 (adenocarcinoma); BG616699 (liver); BE932253 (head\_neck); AL530396 (neuroblastoma cells).

An object of the invention is a purified, isolated, or recombinant nucleic acid comprising the nucleotide sequence of SEQ ID NO: 160, complementary sequences thereto, and fragments thereof.  
35 The invention also pertains to a purified or isolated nucleic acid comprising a polynucleotide having at least 95% nucleotide identity with a polynucleotide of SEQ ID NO: 160, advantageously 99 % nucleotide identity, preferably 99.5% nucleotide identity and most preferably 99.8% nucleotide



identity with a polynucleotide of SEQ ID NO: 160, or a sequence complementary thereto or a biologically active fragment thereof. Another object of the invention relates to purified, isolated or recombinant nucleic acids comprising a polynucleotide that hybridizes, under the stringent hybridization conditions defined herein, with a polynucleotide of SEQ ID NO: 160, or a sequence complementary thereto or a variant thereof or a biologically active fragment thereof. In further  
5 embodiments, nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID NO: 160, or the complements thereof.

Also encompassed is a purified, isolated, or recombinant nucleic acid polynucleotide  
10 encoding a THAP1 polypeptide of the invention, as further described herein.

In another preferred aspect, the invention pertains to purified or isolated nucleic acid molecules that encode a portion or variant of a THAP1 protein, wherein the portion or variant displays a THAP1 activity of the invention. Preferably said portion or variant is a portion or variant of a naturally occurring full-length THAP1 protein. In one example, the invention provides a  
15 polynucleotide comprising, consisting essentially of, or consisting of a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID NO: 160, wherein said nucleic acid encodes a THAP1 portion or variant having a THAP1 activity described herein. In other embodiments, the invention relates to a polynucleotide encoding a THAP1 portion consisting of 8-20, 20-50, 50-70, 60-100, 100 - 150, 150- 200, 200-205 or 205-212  
20 amino acids of SEQ ID NO: 3, or a variant thereof, wherein said THAP1 portion displays a THAP1 activity described herein.

The sequence of SEQ ID NO: 160 corresponds to the human THAP1 cDNA. This cDNA comprises sequences encoding the human THAP1 protein (i.e., "the coding region", from nucleotides 202 to 840, as well as 5' untranslated sequences (nucleotides 1-201) and 3' untranslated  
25 sequences (nucleotides 841 to 2173).

Also encompassed by the THAP1 nucleic acids of the invention are nucleic acid molecules which are complementary to THAP1 nucleic acids described herein. Preferably, a complementary nucleic acid is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO: 160, such that it can hybridize to the nucleotide sequence shown in SEQ ID NO: 160, thereby forming a  
30 stable duplex.

Another object of the invention is a purified, isolated, or recombinant nucleic acid encoding a THAP1 polypeptide comprising, consisting essentially of, or consisting of the amino acid sequence of SEQ ID NO: 3, or fragments thereof, wherein the isolated nucleic acid molecule encodes one or more motifs selected from the group consisting of a THAP domain, a THAP1 target binding region,  
35 a nuclear localization signal and a interferon gamma homology motif. Preferably said THAP1 target binding region is a PAR4 binding region or a DNA binding region. For example, the purified, isolated or recombinant nucleic acid may comprise a genomic DNA or fragment thereof

which encodes the polypeptide of SEQ ID NO: 3 or a fragment thereof or a cDNA consisting of, consisting essentially of, or comprising the sequence of SEQ ID NO: 160 or fragments thereof, wherein the isolated nucleic acid molecule encodes one or more motifs selected from the group consisting of a THAP domain, a THAP1-target binding region, a nuclear localization signal and a  
5 interferon gamma homology motif. Any combination of said motifs may also be specified. Preferably said THAP1 target binding region is a PAR4 binding region or a DNA binding region. Particularly preferred nucleic acids of the invention include isolated, purified, or recombinant THAP1 nucleic acids comprising, consisting essentially of, or consisting of a contiguous span of at  
10 least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or 300 nucleotides of a sequence selected from the group consisting of nucleotide positions ranges consisting of 607 to 708, 637 to 696 and 703 to 747 of SEQ ID NO: 160. In preferred embodiments, a THAP1 nucleic acid encodes a THAP1 polypeptide comprising at least two THAP1 functional domains, such as for example a THAP domain and a PAR4 binding region.

In further preferred embodiments, a THAP1 nucleic acid comprises a nucleotide sequence  
15 encoding a THAP domain having the consensus amino acid sequence of the formula of SEQ ID NOs: 1-2. A THAP1 nucleic acid may also encode a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids to the THAP domain consensus sequence (SEQ ID NOs: 1-2). The present invention also embodies isolated, purified, and recombinant  
20 polynucleotides which encode a polypeptide comprising a contiguous span of at least 6 amino acids, preferably at least 8 or 10 amino acids, more preferably at least 15, 25, 30, 35, 40, 45, 50, 60, 70, 80 or 90 amino acids according to the formula of SEQ ID NO: 1-2.

The nucleotide sequence determined from the cloning of the THAP1 gene allows for the generation of probes and primers designed for use in identifying and/or cloning other THAP1  
25 family members (e.g. sharing the novel functional domains), as well as THAP1 homologues from other species.

A nucleic acid fragment encoding a "biologically active portion of a THAP1 protein" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO: 160, which encodes a polypeptide having a THAP1 biological activity (the biological activities of the THAP1 proteins  
30 described herein), expressing the encoded portion of the THAP1 protein (e.g., by recombinant expression in vitro or in vivo) and assessing the activity of the encoded portion of the THAP1 protein.

The invention further encompasses nucleic acid molecules that differ from the THAP1 nucleotide sequences of the invention due to degeneracy of the genetic code and encode the same  
35 THAP1 proteins and fragment of the invention.

In addition to the THAP1 nucleotide sequences described above, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid

sequences of the THAP1 proteins may exist within a population (e.g., the human population). Such genetic polymorphism may exist among individuals within a population due to natural allelic variation. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of a THAP1 gene.

5 Nucleic acid molecules corresponding to natural allelic variants and homologues of the THAP1 nucleic acids of the invention can be isolated based on their homology to the THAP1 nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

10 Probes based on the THAP1 nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g., the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a THAP1 protein, such as by  
15 measuring a level of a THAP1-encoding nucleic acid in a sample of cells from a subject e.g., detecting THAP1 mRNA levels or determining whether a genomic THAP1 gene has been mutated or deleted.

#### **THAP1 Polypeptides**

The term "THAP1 polypeptides" is used herein to embrace all of the proteins and  
20 polypeptides of the present invention. Also forming part of the invention are polypeptides encoded by the polynucleotides of the invention, as well as fusion polypeptides comprising such polypeptides. The invention embodies THAP1 proteins from humans, including isolated or purified THAP1 proteins consisting of, consisting essentially of, or comprising the sequence of SEQ ID NO:  
3.

25 Aspects of the present invention concern the polypeptide encoded by a nucleotide sequence of SEQ ID NO: 160, a complementary sequence thereof or a fragment thereto.

Another aspect of the present invention embodies isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID NO:  
30 3. In other preferred embodiments the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the THAP1 protein sequence. The invention also concerns the polypeptide encoded by the THAP1 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof.

35 One aspect of the invention pertains to isolated THAP1 proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-THAP1 antibodies. In one embodiment, native THAP1 proteins can be isolated from cells or tissue

sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, THAP1 proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a THAP1 protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

5 Typically, biologically active portions comprise a domain or motif with at least one activity of the THAP1 protein. The present invention also embodies isolated, purified, and recombinant portions or fragments of one THAP1 polypeptide comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 100 or 200 amino acids of SEQ ID NO: 3. Also encompassed are THAP1 polypeptide which comprise  
10 between 10 and 20, between 20 and 50, between 30 and 60, between 50 and 100, or between 100 and 200 amino acids of SEQ ID NO: 3. In other preferred embodiments the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the THAP1 protein sequence.

A biologically active THAP1 protein may, for example, comprise at least 1, 2, 3, 5, 10, 20 or  
15 30 amino acid changes from the sequence of SEQ ID NO: 3, or may encode a biologically active THAP1 protein comprising at least 1%, 2%, 3%, 5%, 8%, 10% or 15% changes in amino acids from the sequence of SEQ ID NO: 3.

In a preferred embodiment, the THAP1 protein comprises, consists essentially of, or consists of a THAP domain at amino acid positions 1 to 89 shown in SEQ ID NO: 3, or fragments or  
20 variants thereof. In other aspects, a THAP1 polypeptide comprises a THAP1-target binding region, a nuclear localization signal and/or a Interferon Gamma Homology Motif. Preferably a THAP1 target binding region is a PAR4 binding region or a DNA binding region. The invention also concerns the polypeptide encoded by the THAP1 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies  
25 isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80, 90 or 100 amino acids of an amino acid sequence selected from the group consisting of positions 1 to 90, 136 to 169, 146 to 165 and 168 to 175 of SEQ ID NO: 3. In another aspect, a THAP1 polypeptide may encode a THAP domain  
30 wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids to the THAP domain consensus sequence (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP1 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 90 shown in SEQ ID NO: 3, or  
35 fragments or variants thereof.

In other embodiments, the THAP1 protein is substantially homologous to the sequences of SEQ ID NO: 3, and retains the functional activity of the THAP1 protein, yet differs in amino acid

sequence due to natural allelic variation or mutagenesis, as described further herein. Accordingly, in another embodiment, the THAP1 protein is a protein which comprises an amino acid sequence shares more than about 60% but less than 100% homology with the amino acid sequence of SEQ ID NO: 3 and retains the functional activity of the THAP1 proteins of SEQ ID NO: 3, respectively.

- 5 Preferably, the protein is at least about 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 92%, 95%, 97%, 98%, 99% or 99.8% homologous to SEQ ID NO: 3, but is not identical to SEQ ID NO: 3. Preferably the THAP1 is less than identical (e.g. 100% identity) to a naturally occurring THAP1. Percent homology can be determined as further detailed above.

#### **THAP-2 to THAP11 and THAP-0 Nucleic Acids**

- 10 As mentioned, the invention provides several members of the THAP-family. THAP-2, THAP-3, THAP-4, THAP-5, THAP-6, THAP-7, THAP-8, THAP-9, THAP10, THAP11 and THAP-0 are described herein. The human and mouse nucleotide sequences corresponding to the human cDNA sequences are listed in SEQ ID NOs: 161-171; and the human amino acid sequences are listed respectively in SEQ ID NOs: 4-14. Also encompassed by the invention are orthologs of  
15 said THAP-family sequences, including mouse, rat, pig and other orthologs, the amino acid sequences of which are listed in SEQ ID NOs: 16-114 and the cDNA sequences are listed in SEQ ID NOs: 172-175.

#### *THAP-2*

- The human THAP-2 cDNA, which is approximately 1302 nucleotides in length shown in  
20 SEQ ID NO: 161, encodes a protein which is approximately 228 amino acid residues in length, shown in SEQ ID NO: 4. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP-2 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The human THAP-2  
25 gene is localized at chromosomes 12 and 3. The THAP-2 protein comprises a THAP domain at amino acids 1 to 89. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP-2 is expressed as follows: BG677995 (squamous cell carcinoma); AV718199 (hypothalamus); BI600215 (hypothalamus); AI208780 (Soares\_testis\_NHT); BE566995  
30 (carcinoma cell line); AI660418 (thymus pooled)

#### *THAP-3*

- The human THAP-3 cDNA which is approximately 1995 nucleotides in length shown in  
SEQ ID NO: 162. The THAP-3 gene encodes a protein which is approximately 239 amino acid residues in length, shown in SEQ ID NO: 5. One aspect of the invention pertains to purified or  
35 isolated nucleic acid molecules that encode THAP-3 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The

human THAP-3 gene is localized at chromosome 1. The THAP-3 protein comprises a THAP domain at amino acids 1 to 89. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP-3 is expressed as follows: BG700517 (hippocampus); BI460812 (testis); BG707197 (hypothalamus); AW960428 (-); BG437177 (large cell carcinoma); BE962820 (adenocarcinoma); BE548411 (cervical carcinoma cell line); AL522189 (neuroblastoma cells); BE545497 (cervical carcinoma cell line); BE280538 (choriocarcinoma); BI086954 (cervix); BE744363 (adenocarcinoma cell line); and BI549151 (hippocampus).

#### THAP-4

The human THAP-4 cDNA, shown as a sequence having 1999 nucleotides in length shown in SEQ ID NO: 163, encodes a protein which is approximately 577 amino acid residues in length, shown in SEQ ID NO: 6. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP-4 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The THAP-4 protein comprises a THAP domain at amino acids 1 to 90. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP-4 is expressed as follows: AL544881 (placenta); BE384014 (melanotic melanoma); AL517205 (neuroblastoma cells); BG394703 (retinoblastoma); BG472327 (retinoblastoma); BI196071 (neuroblastoma); BE255202 (retinoblastoma); BI017349 (lung\_tumor); BF972153 (leiomyosarcoma cell line); BG116061 (duodenal adenocarcinoma cell line); AL530558 (neuroblastoma cells); AL520036 (neuroblastoma cells); AL559902 (B cells from Burkitt lymphoma); AL534539 (Fetal brain); BF686560 (leiomyosarcoma cell line); BF345413 (anaplastic oligodendroglioma with 1p/19q loss); BG117228 (adenocarcinoma cell line); BG490646 (large cell carcinoma); and BF769104 (epid\_tumor).

#### THAP-5

The human THAP-5 cDNA, shown as a sequence having 1034 nucleotides in length shown in SEQ ID NO: 164, encodes a protein which is approximately 239 amino acid residues in length, shown in SEQ ID NO: 7. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP-5 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The human THAP-5 gene is localized at chromosome 7. The THAP-5 protein comprises a THAP domain at amino acids 1 to 90. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP-5 is expressed as follows: BG575430 (mammary adenocarcinoma cell line); BI545812 (hippocampus);

BI560073 (testis); BG530461 (embryonal carcinoma); BF244164 (glioblastoma); BI461364 (testis); AW407519 (germinal center B cells); BF103690 (embryonal carcinoma); and BF939577 (kidney).

#### THAP-6

The human THAP-6 cDNA, shown as a sequence having 2291 nucleotides in length shown in SEQ ID NO: 165, encodes a protein which is approximately 222 amino acid residues in length, shown in SEQ ID NO: 8. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP-6 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The human THAP-6 gene is localized at chromosome 4. The THAP-6 protein comprises a THAP domain at amino acids 1 to 90. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP-6 is expressed as follows: AV684783 (hepatocellular carcinoma); AV698391 (hepatocellular carcinoma); BI560555 (testis); AV688768 (hepatocellular carcinoma); AV692405 (hepatocellular carcinoma); and AV696360 (hepatocellular carcinoma).

#### THAP-7

The human THAP-7 cDNA, shown as a sequence having 1242 nucleotides in length shown in SEQ ID NO: 166, encodes a protein which is approximately 309 amino acid residues in length, shown in SEQ ID NO: 9. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP-7 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The human THAP-7 gene is localized at chromosome 22q11.2. The THAP-7 protein comprises a THAP domain at amino acids 1 to 90. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP-7 is expressed as follows: BI193682 (epithelioid carcinoma cell line); BE253146 (retinoblastoma); BE622113 (melanotic melanoma); BE740360 (adenocarcinoma cell line); BE513955 (Burkitt lymphoma); AL049117 (testis); BF952983 (nervous\_normal); AW975614 (-); BE273270 (renal cell adenocarcinoma); BE738428 (glioblastoma); BE388215 (endometrium adenocarcinoma cell line); BF762401 (colon\_est); and BG329264 (retinoblastoma).

#### THAP-8

The human THAP-8 cDNA, shown as a sequence having 1383 nucleotides in length shown in SEQ ID NO: 167, encodes a protein which is approximately 274 amino acid residues in length, shown in SEQ ID NO: 10. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP-8 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The human THAP-8

gene is localized at chromosome 19. The THAP-8 protein comprises a THAP domain at amino acids 1 to 92. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP-8 is expressed as follows: BG703645 (hippocampus); BF026346 (melanotic melanoma);  
5 BE728495 (melanotic melanoma); BG334298 (melanotic melanoma); and BE390697 (endometrium adenocarcinoma cell line).

#### THAP-9

The human THAP-9 cDNA, shown as a sequence having 693 nucleotides in length shown in SEQ ID NO: 168, encodes a protein which is approximately 231 amino acid residues in length,  
10 shown in SEQ ID NO: 11. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP-9 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The THAP-9 protein comprises a THAP domain at amino acids 1 to 92. Analysis of expressed sequences (accession  
15 numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP-9 is expressed as follows: AA333595 (Embryo 8 weeks).

#### THAP10

The human THAP10 cDNA, shown as a sequence having 771 nucleotides in length shown in SEQ ID NO: 169, encodes a protein which is approximately 257 amino acid residues in length,  
20 shown in SEQ ID NO: 12. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP10 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The human THAP10  
25 gene is localized at chromosome 15. The THAP10 protein comprises a THAP domain at amino acids 1 to 90. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP10 is expressed as follows: AL526710 (neuroblastoma cells); AV725499 (Hypothalamus); AW966404 (-); AW296810 (lung); and AL557817 (T cells from T cell leukemia).

#### THAP11

The human THAP11 cDNA, shown as a sequence having 942 nucleotides in length shown in SEQ ID NO: 170, encodes a protein which is approximately 314 amino acid residues in length,  
35 shown in SEQ ID NO: 13. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP11 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The human THAP11 gene is localized at chromosome 16. The THAP11 protein comprises a THAP domain at amino



acids 1 to 90. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP11 is expressed as follows: AU142300 (retinoblastoma); BI261822 (lymphoma cell line); BG423102 (renal cell adenocarcinoma); and BG423864 (kidney).

5            *THAP-0*

The human THAP-0 cDNA, shown as a sequence having 2283 nucleotides in length shown in SEQ ID NO: 171, encodes a protein which is approximately 761 amino acid residues in length, shown in SEQ ID NO: 14. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP-0 proteins or biologically active portions thereof as further described  
10    herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The human THAP-0 gene is localized at chromosome 11. The THAP-0 protein comprises a THAP domain at amino acids 1 to 90. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that  
15    THAP-0 is expressed as follows: BE713222 (head\_neck); BE161184 (head\_neck); AL119452 (amygdala); AU129709 (teratocarcinoma); AW965460 (-); AW965460(-); AW958065 (-); and BE886885 (leiomyosarcoma).

An object of the invention is a purified, isolated, or recombinant nucleic acid comprising the nucleotide sequence of SEQ ID NOs: 161-171, 173-175 or complementary sequences thereto,  
20    and fragments thereof. The invention also pertains to a purified or isolated nucleic acid comprising a polynucleotide having at least 95% nucleotide identity with a polynucleotide of SEQ ID NOs: 161-171 or 173-175, advantageously 99 % nucleotide identity, preferably 99.5% nucleotide identity and most preferably 99.8% nucleotide identity with a polynucleotide of SEQ ID NOs: 161-171, 173-175 or a sequence complementary thereto or a biologically active fragment thereof. Another  
25    object of the invention relates to purified, isolated or recombinant nucleic acids comprising a polynucleotide that hybridizes, under the stringent hybridization conditions defined herein, with a polynucleotide of SEQ ID NOs: 161-171, 173-175 or a sequence complementary thereto or a variant thereof or a biologically active fragment thereof. In further embodiments, nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous  
30    span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of a sequence selected from the group consisting of SEQ ID NOs: 161-171, 173-175 or the complements thereof.

Also encompassed is a purified, isolated, or recombinant nucleic acid polynucleotide encoding a THAP-2 to THAP11 or THAP-0 polypeptide of the invention, as further described  
35    herein.

In another preferred aspect, the invention pertains to purified or isolated nucleic acid molecules that encode a portion or variant of a THAP-2 to THAP11 or THAP-0 protein, wherein

the portion or variant displays a THAP-2 to THAP11 or THAP-0 activity of the invention. Preferably said portion or variant is a portion or variant of a naturally occurring full-length THAP-2 to THAP11 or THAP-0 protein. In one example, the invention provides a polynucleotide comprising, consisting essentially of, or consisting of a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides, to the extent that the length of said span is consistent with the length of the SEQ ID NO, of a sequence selected from the group consisting of SEQ ID NOs: 161-171, 173-175, wherein said nucleic acid encodes a THAP-2 to THAP11 or THAP-0 portion or variant having a THAP-2 to THAP11 or THAP-0 activity described herein. In other embodiment, the invention relates to a polynucleotide encoding a THAP-2 to THAP11 or THAP-0 portion consisting of 8-20, 20-50, 50-70, 60-100, 100 - 150, 150- 200, 200-250 or 250 - 350 amino acids, to the extent that the length of said portion is consistent with the length of the SEQ ID NO: of a sequence selected from the group consisting of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98, 100-114 or a variant thereof, wherein said THAP-2 to THAP11 or THAP-0 portion displays a THAP-2 to THAP11 or THAP-0 activity described herein.

A THAP-2 to THAP11 or THAP-0 variant nucleic acid may, for example, encode a biologically active THAP-2 to THAP11 or THAP-0 protein comprising at least 1, 2, 3, 5, 10, 20 or 30 amino acid changes from the respective sequence selected from the group consisting of SEQ ID NO: 4-14, 17-21, 23-40, 42-56, 58-98 and 100-114 or may encode a biologically active THAP-2 to THAP11 or THAP-0 protein comprising at least 1%, 2%, 3%, 5%, 8%, 10% or 15% changes in amino acids from the respective sequence of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 and 100-114.

The sequences of SEQ ID NOs: 4-14 correspond to the human THAP-2 to THAP11 and THAP-0 DNAs respectively. SEQ ID NOs: 17-21, 23-40, 42-56, 58-98, 100-114 correspond to mouse, rat, pig and other orthologs.

Also encompassed by the THAP-2 to THAP11 and THAP-0 nucleic acids of the invention are nucleic acid molecules which are complementary to THAP-2 to THAP11 or THAP-0 nucleic acids described herein. Preferably, a complementary nucleic acid is sufficiently complementary to the nucleotide respective sequence shown in SEQ ID NOs: 161-171 and 173-175 such that it can hybridize to said nucleotide sequence shown in SEQ ID NOs: 161-171 and 173-175, thereby forming a stable duplex.

Another object of the invention is a purified, isolated, or recombinant nucleic acid encoding a THAP-2 to THAP11 or THAP-0 polypeptide comprising, consisting essentially of, or consisting of an amino acid sequence selected from the group consisting of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98, 100-114 or fragments thereof, wherein the isolated nucleic acid molecule encodes a THAP domain or a THAP-2 to THAP11 or THAP-0 target binding region. Preferably said target binding region is a protein binding region, preferably a PAR-4 binding region, or preferably said target binding region is a DNA binding region. For example, the purified, isolated or recombinant

nucleic acid may comprise a genomic DNA or fragment thereof which encodes a polypeptide having a sequence selected from the group consisting of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98, 100-114 or a fragment thereof. The purified, isolated or recombinant nucleic acid may alternatively comprise a cDNA consisting of, consisting essentially of, or comprising a sequence  
5 selected from the group consisting of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98, 100-114 or fragments thereof, wherein the isolated nucleic acid molecule encodes a THAP domain or a THAP-2 to THAP11 or THAP-0 target binding region. In preferred embodiments, a THAP-2 to THAP11 or THAP-0 nucleic acid encodes a THAP-2 to THAP11 or THAP-0 polypeptide comprising at least two THAP-2 to THAP11 or THAP-0 functional domains, such as for example a THAP domain and  
10 a THAP-2 to THAP11 or THAP-0 target binding region.

Particularly preferred nucleic acids of the invention include isolated, purified, or recombinant THAP-2 to THAP11 or THAP-0 nucleic acids comprising, consisting essentially of, or consisting of a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or 250 nucleotides of a sequence selected from the group consisting of nucleotide  
15 positions coding for the relevant amino acids as given in the SEQ ID NO: 161-171 and 173-175.

In further preferred embodiments, a THAP-2 to THAP11 or THAP-0 nucleic acid comprises a nucleotide sequence encoding a THAP domain having the consensus amino acid sequence of the formula of SEQ ID NOs: 1-2. A THAP-2 to THAP11 or THAP-0 nucleic acid may also encode a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least  
20 about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids to the THAP consensus domain (SEQ ID NOs: 1-2). The present invention also embodies isolated, purified, and recombinant polynucleotides which encode a polypeptide comprising a contiguous span of at least 6 amino acids, preferably at least 8 or 10 amino acids, more preferably at least 15, 25, 30, 35, 40, 45, 50, 60, 70, 80 or 90 amino acids of SEQ ID NOs: 1-2

25 .  
The nucleotide sequence determined from the cloning of the THAP-2 to THAP11 or THAP-0 genes allows for the generation of probes and primers designed for use in identifying and/or cloning other THAP family members, particularly sequences related to THAP-2 to THAP11 or THAP-0 (e.g. sharing the novel functional domains), as well as THAP-2 to THAP11 or THAP-0  
30 homologues from other species.

A nucleic acid fragment encoding a biologically active portion of a THAP-2 to THAP11 or THAP-0 protein can be prepared by isolating a portion of a nucleotide sequence selected from the group consisting of SEQ ID NOs: 161-171 and 173-175, which encodes a polypeptide having a THAP-2 to THAP11 or THAP-0 biological activity (the biological activities of the THAP-family  
35 proteins described herein), expressing the encoded portion of the THAP-2 to THAP11 or THAP-0 protein (e.g., by recombinant expression in vitro or in vivo) and assessing the activity of the encoded portion of the THAP-2 to THAP11 or THAP-0 protein.

The invention further encompasses nucleic acid molecules that differ from the THAP-2 to THAP11 or THAP-0 nucleotide sequences of the invention due to degeneracy of the genetic code and encode the same THAP-2 to THAP11 or THAP-0 protein, or fragment thereof, of the invention.

In addition to the THAP-2 to THAP11 or THAP-0 nucleotide sequences described above, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the respective THAP-2 to THAP11 or THAP-0 protein may exist within a population (e.g., the human population). Such genetic polymorphism may exist among individuals within a population due to natural allelic variation. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of a particular THAP-2 to THAP11 or THAP-0 gene.

Nucleic acid molecules corresponding to natural allelic variants and homologues of the THAP-2 to THAP11 or THAP-0 nucleic acids of the invention can be isolated based on their homology to the THAP-2 to THAP11 or THAP-0 nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Probes based on the THAP-2 to THAP11 or THAP-0 nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g., the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a THAP-2 to THAP11 or THAP-0 protein, such as by measuring a level of a THAP-2 to THAP11 or THAP-0-encoding nucleic acid in a sample of cells from a subject e.g., detecting THAP-2 to THAP11 or THAP-0 mRNA levels or determining whether a genomic THAP-2 to THAP11 or THAP-0 gene has been mutated or deleted.

#### **THAP-2 to THAP11 and THAP-0 Polypeptides**

The term "THAP-2 to THAP11 or THAP-0 polypeptides" is used herein to embrace all of the proteins and polypeptides of the present invention relating to THAP-2, THAP-3, THAP-4, THAP-5, THAP-6, THAP-7, THAP-8, THAP-9, THAP10, THAP11 and THAP-0. Also forming part of the invention are polypeptides encoded by the polynucleotides of the invention, as well as fusion polypeptides comprising such polypeptides. The invention embodies THAP-2 to THAP11 or THAP-0 proteins from humans, including isolated or purified THAP-2 to THAP11 or THAP-0 proteins consisting of, consisting essentially of, or comprising a sequence selected from the group consisting of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 and 100-114.

The invention concerns the polypeptide encoded by a nucleotide sequence selected from the group consisting of SEQ ID NOs: 161-171, 172-175 and a complementary sequence thereof and a fragment thereof.

The present invention embodies isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 100, 150, 200, 300 or 500 amino acids, to the extent that said span is consistent with the particular SEQ ID NO:, of a sequence selected from the group consisting of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 and 100-114. In other preferred  
5      embodiments the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the THAP-2 to THAP11 or THAP-0 protein sequence.

One aspect of the invention pertains to isolated THAP-2 to THAP11 and THAP-0 proteins,  
10     and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-THAP-2 to THAP11 or THAP-0 antibodies. In one embodiment, native THAP-2 to THAP11 or THAP-0 proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, THAP-2 to THAP11 or THAP-0 proteins are produced by recombinant DNA  
15     techniques. Alternative to recombinant expression, a THAP-2 to THAP11 or THAP-0 protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

Biologically active portions of a THAP-2 to THAP11 or THAP-0 protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the THAP-2 to THAP11 or THAP-0 protein, e.g., an amino acid sequence shown in  
20     SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 or 100-114, which include less amino acids than the respective full length THAP-2 to THAP11 or THAP-0 protein, and exhibit at least one activity of the THAP-2 to THAP11 or THAP-0 protein. The present invention also embodies isolated, purified, and recombinant portions or fragments of a THAP-2 to THAP11 or THAP-0 polypeptide comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids,  
25     more preferably at least 12, 15, 20, 25, 30, 40, 50, 100, 150, 200, 300 or 500 amino acids, to the extent that said span is consistent with the particular SEQ ID NO, of a sequence selected from the group consisting of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 and 100-114. Also encompassed are THAP-2 to THAP11 or THAP-0 polypeptides which comprise between 10 and 20, between 20 and 50, between 30 and 60, between 50 and 100, or between 100 and 200 amino  
30     acids of a sequence selected from the group consisting of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 and 100-114. In other preferred embodiments the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the THAP-2 to THAP11 or THAP-0 protein sequence.

A biologically active THAP-2 to THAP11 or THAP-0 protein may, for example, comprise at  
35     least 1, 2, 3, 5, 10, 20 or 30 amino acid changes from the sequence of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 or 100-114, or may encode a biologically active THAP-2 to THAP11 or

THAP-0 protein comprising at least 1%, 2%, 3%, 5%, 8%, 10% or 15% changes in amino acids from the sequence of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 or 100-114.

In a preferred embodiment, the THAP-2 protein comprises, consists essentially of, or consists of a THAP-2 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 89 shown in SEQ ID NO: 4, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP-2 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 89 amino acids of a sequence comprising amino acid positions 1 to 89 of SEQ ID NO: 4. In another aspect, a THAP-2 polypeptide may comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids-to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP-2 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 89 shown in SEQ ID NO: 4, or fragments or variants thereof. Preferably, said THAP-2 polypeptide comprises a PAR-4 binding domain and/or a DNA binding domain.

In a preferred embodiment, the THAP-3 protein comprises, consists essentially of, or consists of a THAP-3 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 89 shown in SEQ ID NO: 5, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP-3 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 89 amino acids of a sequence comprising amino acid positions 1 to 89 of SEQ ID NO: 5. In another aspect, a THAP-3 polypeptide may comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids-to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP-3 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 89 shown in SEQ ID NO: 5, or fragments or variants thereof. Preferably, said THAP-3 polypeptide comprises a PAR-4 binding domain and/or a DNA binding domain.

In a preferred embodiment, the THAP-4 protein comprises, consists essentially of, or consists of a THAP-4 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 90 shown in SEQ ID NO: 6, or fragments or variants thereof. The invention also concerns the

polypeptide encoded by the THAP-4 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 90 amino acids of a sequence comprising amino acid positions 1 to 90 of SEQ ID NO: 6. In another aspect, a THAP-4 polypeptide may comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP-4 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 90 shown in SEQ ID NO: 6, or fragments or variants thereof.

In a preferred embodiment, the THAP-5 protein comprises, consists essentially of, or consists of a THAP-5 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 90 shown in SEQ ID NO: 7, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP-5 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 90 amino acids of a sequence comprising amino acid positions 1 to 90 of SEQ ID NO: 7. In another aspect, a THAP-5 polypeptide may comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP-5 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 90 shown in SEQ ID NO: 7, or fragments or variants thereof.

In a preferred embodiment, the THAP-6 protein comprises, consists essentially of, or consists of a THAP-6 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 90 shown in SEQ ID NO: 8, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP-6 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 90 amino acids of a sequence comprising amino acid positions 1 to 90 of SEQ ID NO: 8. In another aspect, a THAP-6 polypeptide may comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably

at least about 65% of the amino acid residues are identical or similar amino acids-to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP-6 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 90 shown in SEQ ID NO: 8, or  
5 fragments or variants thereof.

In a preferred embodiment, the THAP-7 protein comprises, consists essentially of, or consists of a THAP-7 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 90 shown in SEQ ID NO: 9, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP-7 nucleotide sequences of the invention, or a complementary  
10 sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 90 amino acids of a sequence comprising amino acid positions 1 to 90 of SEQ ID NO: 9. In another aspect, a THAP-7 polypeptide may comprise a THAP domain  
15 wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids-to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP-7 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 90 shown in SEQ ID NO: 9, or  
20 fragments or variants thereof.

In a preferred embodiment, the THAP-8 protein comprises, consists essentially of, or consists of a THAP-8 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 92 shown in SEQ ID NO: 10, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP-8 nucleotide sequences of the invention, or a complementary  
25 sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 90 amino acids of a sequence comprising amino acid positions 1 to 92 of SEQ ID NO: 10. In another aspect, a THAP-8 polypeptide may comprise a THAP domain  
30 wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids-to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP-8 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 92 shown in SEQ ID NO: 10, or  
35 fragments or variants thereof.

In a preferred embodiment, the THAP-9 protein comprises, consists essentially of, or consists of a THAP-9 THAP domain, preferably having the amino acid sequence of amino acid



positions 1 to 92 shown in SEQ ID NO: 11, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP-9 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 90 amino acids of a sequence comprising amino acid positions 1 to 92 of SEQ ID NO: 11. In another aspect, a THAP-9 polypeptide may comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids-to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP-9 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 92 shown in SEQ ID NO: 11, or fragments or variants thereof.

In a preferred embodiment, the THAP10 protein comprises, consists essentially of, or consists of a THAP10 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 90 shown in SEQ ID NO: 12, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP10 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 90 amino acids of a sequence comprising amino acid positions 1 to 90 of SEQ ID NO: 12. In another aspect, a THAP10 polypeptide may comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids-to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP10 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 90 shown in SEQ ID NO: 12, or fragments or variants thereof.

In a preferred embodiment, the THAP11 protein comprises, consists essentially of, or consists of a THAP11 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 90 shown in SEQ ID NO: 13, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP11 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 90 amino acids of a sequence comprising amino acid positions 1 to 90 of SEQ ID NO: 13. In another aspect, a THAP11 polypeptide may

comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids-to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP11 polypeptide comprising,  
5 consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 90 shown in SEQ ID NO: 13, or fragments or variants thereof.

In a preferred embodiment, the THAP-0 protein comprises, consists essentially of, or consists of a THAP-0 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 90 shown in SEQ ID NO: 14, or fragments or variants thereof. The invention also  
10 concerns the polypeptide encoded by the THAP-0 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 90 amino acids of a sequence comprising  
15 amino acid positions 1 to 90 of SEQ ID NO: 14. In another aspect, a THAP-0 polypeptide may comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids-to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP-0 polypeptide comprising,  
20 consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 90 shown in SEQ ID NO: 14, or fragments or variants thereof.

In other embodiments, the THAP-2 to THAP11 or THAP-0 protein is substantially homologous to the sequences of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 or 100-114 and retains the functional activity of the THAP-2 to THAP11 or THAP-0 protein, yet differs in amino  
25 acid sequence due to natural allelic variation or mutagenesis, as described further herein. Accordingly, in another embodiment, the THAP-2 to THAP11 or THAP-0 protein is a protein which comprises an amino acid sequence that shares more than about 60% but less than 100% homology with the amino acid sequence of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 or 100-114 and retains the functional activity of the THAP-2 to THAP11 or THAP-0 proteins of SEQ ID  
30 NOs: 4-14, 17-21, 23-40, 42-56, 58-98 or 100-114, respectively. Preferably, the protein is at least about 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 92%, 95%, 97%, 98%, 99% or 99.8% homologous to SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 or 100-114, but is not identical to SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 or 100-114. Preferably the THAP-2 to THAP11 or THAP-0 is less than identical (e.g. 100% identity) to a naturally occurring THAP-2 to THAP11 or  
35 THAP-0. Percent homology can be determined as further detailed above.

**Assessing polypeptides, methods for obtaining variant nucleic acids and polypeptides**

It will be appreciated that by characterizing the function of THAP-family polypeptides, the invention further provides methods of testing the activity of, or obtaining, functional fragments and variants of THAP-family and THAP domain nucleotide sequences involving providing a variant or modified THAP-family or THAP domain nucleic acid and assessing whether a polypeptide encoded thereby displays a THAP-family activity of the invention. Encompassed is thus a method of assessing the function of a THAP-family or THAP domain polypeptide comprising : (a) providing a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof; and (b) testing said THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof for a THAP-family activity. Any suitable format may be used, including cell free, cell-based and in vivo formats. For example, said assay may comprise expressing a THAP-family or THAP domain nucleic acid in a host cell, and observing THAP-family activity in said cell. In another example, a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof is introduced to a cell, and a THAP-family activity is observed. THAP-family activity may be any activity as described herein, including- (1) mediating apoptosis or cell proliferation when expressed or introduced into a cell, most preferably inducing or enhancing apoptosis, and/or most preferably reducing cell proliferation; (2) mediating apoptosis or cell proliferation of an endothelial cell; (3) mediating apoptosis or cell proliferation of a hyperproliferative cell; (4) mediating apoptosis or cell proliferation of a CNS cell, preferably a neuronal or glial cell; or (5) an activity determined in an animal selected from the group consisting of mediating, preferably inhibiting angiogenesis, mediating, preferably inhibiting inflammation, inhibition of metastatic potential of cancerous tissue, reduction of tumor burden, increase in sensitivity to chemotherapy or radiotherapy, killing a cancer cell, inhibition of the growth of a cancer cell, or induction of tumor regression.

In addition to naturally-occurring allelic variants of the THAP-family or THAP domain sequences that may exist in the population, the skilled artisan will appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NOs: 160-171, thereby leading to changes in the amino acid sequence of the encoded THAP-family or THAP domain proteins, with or without altering the functional ability of the THAP-family or THAP domain proteins.

Several types of variants are contemplated including 1) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue and such substituted amino acid residue may or may not be one encoded by the genetic code, or 2) one in which one or more of the amino acid residues includes a substituent group, or 3) one in which the mutated THAP-family or THAP domain polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or 4) one in which the additional amino acids are fused to the mutated THAP-family or THAP domain polypeptide, such as a leader or secretory sequence or a sequence which is employed for

purification of the mutated THAP-family or THAP domain polypeptide or a preprotein sequence. Such variants are deemed to be within the scope of those skilled in the art.

For example, nucleotide substitutions leading to amino acid substitutions can be made in the sequences of SEQ ID NOs: 160-175 that do not substantially change the biological activity of the protein. An amino acid residue can be altered from the wild-type sequence encoding a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof without altering the biological activity. In general, amino acid residues that are conserved among the THAP-family of THAP domain-containing proteins of the present invention, are predicted to be less amenable to alteration. Furthermore, additional conserved amino acid residues may be amino acids that are conserved between the THAP-family proteins of the present invention.

In one aspect, the invention pertains to nucleic acid molecules encoding THAP family or THAP domain polypeptides, or biologically active fragments or homologues thereof that contain changes in amino acid residues that are not essential for activity. Such THAP-family proteins differ in amino acid sequence from SEQ ID NOs: 1-114 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 60% homologous to an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114. Preferably, the protein encoded by the nucleic acid molecule is at least about 65-70% homologous to an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114, more preferably sharing at least about 75-80% identity with an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114, even more preferably sharing at least about 85%, 90%, 92%, 95%, 97%, 98%, 99% or 99.8% identity with an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114.

In another aspect, the invention pertains to nucleic acid molecules encoding THAP-family proteins that contain changes in amino acid residues that result in increased biological activity, or a modified biological activity. In another aspect, the invention pertains to nucleic acid molecules encoding THAP-family proteins that contain changes in amino acid residues that are essential for a THAP-family activity. Such THAP-family proteins differ in amino acid sequence from SEQ ID NOs: 1-114 and display reduced or essentially lack one or more THAP-family biological activities. The invention also encompasses a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof which may be useful as dominant negative mutant of a THAP family or THAP domain polypeptide.

An isolated nucleic acid molecule encoding a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof homologous to a protein of any one of SEQ ID NOs: 1-114 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NOs: 1-114 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be

introduced into any of SEQ ID NOs: 1-114, by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. For example, conservative amino acid substitutions may be made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue  
5 having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched  
10 side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof may be replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a THAP-family or  
15 THAP domain coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for THAP-family biological activity to identify mutants that retain activity. Following mutagenesis of one of SEQ ID NOs: 1-114, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant THAP family or THAP domain polypeptide, or a  
20 biologically active fragment or homologue thereof encoded by a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof of THAP domain nucleic acid of the invention can be assayed for a THAP-family activity in any suitable assay, examples of which are provided herein.

The invention also provides THAP-family or THAP domain chimeric or fusion proteins. As  
25 used herein, a THAP-family or THAP domain "chimeric protein" or "fusion protein" comprises a THAP-family or THAP domain polypeptide of the invention operatively linked, preferably fused in frame, to a non-THAP-family or non-THAP domain polypeptide. In a preferred embodiment, a THAP-family or THAP domain fusion protein comprises at least one biologically active portion of a THAP-family or THAP domain protein. In another preferred embodiment, a THAP-family fusion  
30 protein comprises at least two biologically active portions of a THAP-family protein. For example, in one embodiment, the fusion protein is a GST-THAP-family fusion protein in which the THAP-family sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant THAP-family polypeptides. In another embodiment, the fusion protein is a THAP-family protein containing a heterologous signal sequence at its N-  
35 terminus, such as for example to allow for a desired cellular localization in a certain host cell.

The THAP-family or THAP domain fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject in vivo. Moreover, the THAP-

family-fusion or THAP domain proteins of the invention can be used as immunogens to produce anti-THAP-family or anti or THAP domain antibodies in a subject, to purify THAP-family or THAP domain ligands and in screening assays to identify molecules which inhibit the interaction of THAP-family or THAP domain with a THAP-family or THAP domain target molecule.

5 Furthermore, isolated peptidyl portions of the subject THAP-family or THAP domain proteins can also be obtained by screening peptides recombinantly produced from the corresponding fragment of the nucleic acid encoding such peptides. In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, a THAP-family or THAP domain protein of the  
10 present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or preferably divided into overlapping fragments of a desired length. The fragments can be produced (recombinantly or by chemical synthesis) and tested to identify those peptidyl fragments which can function as either agonists or antagonists of a THAP-family protein activity, such as by microinjection assays or in vitro protein binding assays. In an illustrative embodiment,  
15 peptidyl portions of a THAP-family protein, such as a THAP domain or a THAP-family target binding region (e.g. PAR4 in the case of THAP1, THAP-2 and THAP-3), can be tested for THAP-family activity by expression as thioredoxin fusion proteins, each of which contains a discrete fragment of the THAP-family protein (see, for example, U.S. Patents 5,270,181 and 5,292,646; and PCT publication W094/02502).

20 The present invention also pertains to variants of the THAP-family or THAP domain proteins which function as either THAP-family or THAP domain mimetics or as THAP-family or THAP domain inhibitors. Variants of the THAP-family or THAP domain proteins can be generated by mutagenesis, e.g., discrete point mutation or truncation of a THAP-family or THAP domain protein. An agonist of a THAP-family or THAP domain protein can retain substantially the same,  
25 or a subset, of the biological activities of the naturally occurring form of a THAP-family or THAP domain protein. An antagonist of a THAP-family or THAP domain protein can inhibit one or more of the activities of the naturally occurring form of the THAP-family or THAP domain protein by, for example, competitively inhibiting the association of a THAP-family or THAP domain protein with a THAP-family target molecule. Thus, specific biological effects can be elicited by treatment  
30 with a variant of limited function. In one embodiment, variants of a THAP-family or THAP domain protein which function as either THAP-family or THAP domain agonists (mimetics) or as THAP-family or THAP domain antagonists can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of a THAP-family or THAP domain protein for THAP-family or THAP domain protein agonist or antagonist activity. In one embodiment, a variegated library of  
35 THAP-family variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of THAP-family variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene

sequences such that a degenerate set of potential THAP-family sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of THAP-family sequences therein. There are a variety of methods which can be used to produce libraries of potential THAP-family variants from a degenerate oligonucleotide  
5 sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential THAP-family sequences.

In addition, libraries of fragments of a THAP-family or THAP domain protein coding  
10 sequence can be used to generate a variegated population of THAP-family or THAP domain fragments for screening and subsequent selection of variants of a THAP-family or THAP domain protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a THAP-family coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded  
15 DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the THAP-family protein.

20 Modified THAP-family or THAP domain proteins can be used for such purposes as enhancing therapeutic or prophylactic efficacy, or stability (e.g., ex vivo shelf life and resistance to proteolytic degradation in vivo). Such modified peptides, when designed to retain at least one activity of the naturally occurring form of the protein, are considered functional equivalents of the THAP-family or THAP domain protein described in more detail herein. Such modified peptide can  
25 be produced, for instance, by amino acid substitution, deletion, or addition.

Whether a change in the amino acid sequence of a peptide results in a functional THAP-family or THAP domain homolog (e.g. functional in the sense that it acts to mimic or antagonize the wild-type form) can be readily determined by assessing the ability of the variant peptide to produce a response in cells in a fashion similar to the wild-type THAP-family or THAP domain  
30 protein or competitively inhibit such a response. Peptides in which more than one replacement has taken place can readily be tested in the same manner.

This invention further contemplates a method of generating sets of combinatorial mutants of the presently disclosed THAP-family or THAP domain proteins, as well as truncation and fragmentation mutants, and is especially useful for identifying potential variant sequences which  
35 are functional in binding to a THAP-family- or THAP domain- target protein but differ from a wild-type form of the protein by, for example, efficacy, potency and/or intracellular half-life. One purpose for screening such combinatorial libraries is, for example, to isolate novel THAP-family or

THAP domain homologs which function as either an agonist or an antagonist of the biological activities of the wild-type protein, or alternatively, possess novel activities all together. For example, mutagenesis can give rise to THAP-family homologs which have intracellular half-lives dramatically different than the corresponding wild-type protein. The altered protein can be rendered either more stable or less stable to proteolytic degradation or other cellular process which result in destruction of, or otherwise inactivation of, a THAP-family protein. Such THAP-family homologs, and the genes which encode them, can be utilized to alter the envelope of expression for a particular recombinant THAP-family protein by modulating the half-life of the recombinant protein. For instance, a short half-life can give rise to more transient biological effects associated with a particular recombinant THAP-family protein and, when part of an inducible expression system, can allow tighter control of recombinant protein levels within a cell. As above, such proteins, and particularly their recombinant nucleic acid constructs, can be used in gene therapy protocols.

In an illustrative embodiment of this method, the amino acid sequences for a population of THAP-family homologs or other related proteins are aligned, preferably to promote the highest homology possible. Such a population of variants can include, for example, THAP-family homologs from one or more species, or THAP-family homologs from the same species but which differ due to mutation. Amino acids which appear at each position of the aligned sequences are selected to create a degenerate set of combinatorial sequences. There are many ways by which the library of potential THAP-family homologs can be generated from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be carried out in an automatic DNA synthesizer, and the synthetic genes then be ligated into an appropriate gene for expression. The purpose of a degenerate set of genes is to provide, in one mixture, all of the sequences encoding the desired set of potential THAP-family sequences. The synthesis of degenerate oligonucleotides is well known in the art (see for example. Narang, SA (1983) Tetrahedron 393; Itakura et al. (1981) Recombinant DNA, Proc 3rd Cleveland Sympos. Macromolecules, ed. AG Walton, Amsterdam: Elsevier pp. 273-289; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al. (1984) Science 198:1056; Ike et al. (1983) Nucleic Acid Res. 11:477. Such techniques have been employed in the directed evolution of other proteins (see, for example, Scott et al. (1990) Science 249:386-390; Roberts et al. (1992) PNAS 89:2429-2433; Devlin et al. (1990) Science 249:404-406; Cwirla et al. (1990) PNAS 87: 6378-6382; as well as U.S. Patents Nos 5,223,409, 5,198,346, and 5,096,815).

Alternatively, other forms of mutagenesis can be utilized to generate a combinatorial library, particularly where no other naturally occurring homologs have yet been sequenced. For example, THAP-family homologs (both agonist and antagonist forms) can be generated and isolated from a library by screening using, for example, alanine scanning mutagenesis and the like (Ruf et al. (1994) Biochemistry 33:1565-1572; Wang et al. (1994) J Biol. Chem. 269:3095-3099;



Balint et al. (1993) *Gene* 137:109-118; Grodberg et al. (1993) *Eur. J Biochem.* 218:597-601; Nagashima et al. (1993) *J Biol. Chem.* 268:2888-2892; Lowman et al. (1991) *Biochemistry* 30:10832-10838; and Cunningham et al. (1989) *Science* 244:1081-1085), by linker scanning mutagenesis (Gustin et al. (1993) *Virology* 193:653-660; Brown et al. (1992) *Mol. Cell Biol.* 12:2644-2652; McKnight et al. (1982) *Science* 232:316); by saturation mutagenesis (Meyers et al. (1986) *Science* 232:613); by PCR mutagenesis (Leung et al. (1989) *Method Cell Mol Biol* 1: 1-19); or by random mutagenesis (Miller et al. (1992) *A Short Course in Bacterial Genetics*, CSHL Press, Cold Spring Harbor, NY; and Greener et al. (1994) *Strategies in Mol Biol* 7:32-34).

A wide range of techniques are known in the art for screening gene products of combinatorial libraries made by point mutations, as well as for screening cDNA libraries for gene products having a certain property. Such techniques will be generally adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of THAP-family proteins. The most widely used techniques for screening large gene libraries typically comprises cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates relatively easy isolation of the vector encoding the gene whose product was detected.

Each of the illustrative assays described below are amenable to high through-put analysis as necessary to screen large numbers of degenerate THAP-family or THAP domain sequences created by combinatorial mutagenesis techniques. In one screening assay, the candidate gene products are displayed on the surface of a cell or viral particle, and the ability of particular cells or viral particles to bind a THAP-family target molecule (protein or DNA) via this gene product is detected in a "panning assay". For instance, the gene library can be cloned into the gene for a surface membrane protein of a bacterial cell, and the resulting fusion protein detected by panning (Ladner et al., WO 88/06630; Fuchs et al. (1991) *BiolTechnology* 9:1370-1371, and Goward et al. (1992) *TIBS* 18:136-140). In a similar fashion, fluorescently labeled THAP-family target can be used to score for potentially functional THAP-family homologs. Cells can be visually inspected and separated under a fluorescence microscope, or, where the morphology of the cell permits, separated by a fluorescence-activated cell sorter.

In an alternate embodiment, the gene library is expressed as a fusion protein on the surface of a viral particle. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at very high concentrations, a large number of phage can be screened at one time. Second, since each infectious phage displays the combinatorial gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical *E. coli* filamentous phages M13, fd, and f1 are most often used in phage display libraries, as either of the

phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle (Ladner et al. PCT publication WO 90/02909; Garrard et al., PCT publication WO 92/09690; Marks et al. (1992) J Biol. Chem. 267:16007-16010; Griffiths et al. (1993) EMBO J 12:725-734; Clackson et al. (1991) Nature 352:624-628; and Barbas et al. (1992) PNAS 89:4457-4461). In an illustrative embodiment, the recombinant phage antibody system (RPAS, Pharmacia Catalog number 27-9400-01) can be easily modified for use in expressing THAP-family combinatorial libraries, and the THAP-family phage library can be panned on immobilized THAP family target molecule (glutathione immobilized THAP-family target-GST fusion proteins or immobilized DNA). Successive rounds of phage amplification and panning can greatly enrich for THAP-family homologs which retain an ability to bind a THAP-family target and which can subsequently be screened further for biological activities in automated assays, in order to distinguish between agonists and antagonists.

The invention also provides for identification and reduction to functional minimal size of the THAP-family domains, particularly a THAP domain of the subject THAP-family to generate mimetics, e.g. peptide or non-peptide agents, which are able to disrupt binding of a polypeptide of the present invention with a THAP-family target molecule (protein or DNA). Thus, such mutagenic techniques as described above are also useful to map the determinants of THAP-family proteins which participate in protein-protein or protein-DNA interactions involved in, for example, binding to a THAP-family or THAP domain target protein or DNA. To illustrate, the critical residues of a THAP-family protein which are involved in molecular recognition of the THAP-family target can be determined and used to generate THAP-family target-13P-derived peptidomimetics that competitively inhibit binding of the THAP-family protein to the THAP-family target. By employing, for example, scanning mutagenesis to map the amino acid residues of a particular THAP-family protein involved in binding a THAP-family target, peptidomimetic compounds can be generated which mimic those residues in binding to a THAP-family target, and which, by inhibiting binding of the THAP-family protein to the THAP-family target molecule, can interfere with the function of a THAP-family protein in transcriptional regulation of one or more genes. For instance, non hydrolyzable peptide analogs of such residues can be generated using retro-inverse peptides (e.g., see U.S. Patents 5,116,947 and 5,219,089; and Pallai et al. (1983) Int J Pept Protein Res 21:84-92), benzodiazepine (e.g., see Freidinger et al. in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al. in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), substituted gamma lactam rings (Garvey et al. in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), keto-methylene pseudo-peptides (Ewenson et al. (1986) J Med Chem 29:295; and Ewenson et al. in Peptides: Structure and Function (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985), P-turn dipeptide cores (Nagai et al. (1985) Tetrahedron Lett 26:647; and Sato et al. (1986) J Chem

Soc Perkin Trans 1: 123 1), and P-aminoalcohols (Gordon et al. (1985) Biochem Biophys Res Commun 126:419; and Dann et al. (1986) Biochem Biophys Res Commun 134:71).

An isolated THAP-family or THAP domain protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind THAP-family or THAP domain proteins using standard techniques for polyclonal and monoclonal antibody preparation. A full-length THAP-family protein can be used or, alternatively, the invention provides antigenic peptide fragments of THAP-family or THAP domain proteins for use as immunogens. Any fragment of the THAP-family or THAP domain protein which contains at least one antigenic determinant may be used to generate antibodies. The antigenic peptide of a THAP-family or THAP domain protein comprises at least 8 amino acid residues of an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114 and encompasses an epitope of a THAP-family or THAP domain protein such that an antibody raised against the peptide forms a specific immune complex with a THAP-family or THAP domain protein. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

Preferred epitopes encompassed by the antigenic peptide are regions of a THAP-family or THAP domain protein that are located on the surface of the protein, e.g., hydrophilic regions.

A THAP-family or THAP domain protein immunogen typically is used to prepare antibodies by immunizing a suitable subject, (e.g., rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed THAP-family or THAP domain protein or a chemically synthesized THAP-family or THAP domain polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic THAP-family or THAP domain protein preparation induces a polyclonal anti-THAP-family or THAP domain protein antibody response.

The invention concerns antibody compositions, either polyclonal or monoclonal, capable of selectively binding, or selectively bind to an epitope-containing a polypeptide comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 100, or more than 100 amino acids of an amino acid sequence selected from the group consisting of amino acid positions 1 to approximately 90 of SEQ ID NOs: 1-114. The invention also concerns a purified or isolated antibody capable of specifically binding to a mutated THAP-family or THAP domain protein or to a fragment or variant thereof comprising an epitope of the mutated THAP-family or THAP domain protein.

#### **Oligomeric Forms of THAP1**

Certain embodiments of the present invention encompass THAP1 polypeptides in the form of oligomers, such as dimers, trimers, or higher oligomers. Oligomers may be formed by disulfide bonds between cysteine residues on different THAP1 polypeptides, for example. In other

embodiments, oligomers comprise from two to four THAP1 polypeptides joined by covalent or non-covalent interactions between peptide moieties fused to the THAP1 polypeptides. Such peptide moieties may be peptide linkers (spacers), or peptides that have the property of promoting oligomerization. Leucine zippers and certain polypeptides derived from antibodies are among the peptides that can promote oligomerization of THAP1 polypeptides attached thereto. DNA sequences encoding THAP1 oligomers, or fusion proteins that are components of such oligomers, are provided herein.

In one embodiment of the invention, oligomeric THAP1 may comprise two or more THAP1 polypeptides joined through peptide linkers. Examples include those peptide linkers described in U.S. Patent No. 5,073,627. Fusion proteins comprising multiple THAP1 polypeptides separated by peptide linkers may be produced using conventional recombinant DNA technology.

Another method for preparing THAP1 oligomers involves use of a leucine zipper. Leucine zipper domains are peptides that promote oligomerization of the proteins in which they are found. Leucine zippers were originally identified in several DNA-binding proteins (Landschulz et al., Science 240:1759, 1988), and have since been found in a variety of different proteins. Among the known leucine zippers are naturally occurring peptides and derivatives thereof that dimerize or trimerize. Examples of leucine zipper domains suitable for producing THAP1 oligomers are those described International Publication WO 94/10308. Recombinant fusion proteins comprising a THAP1 polypeptide fused to a peptide that dimerizes or trimerizes in solution are expressed in suitable host cells, and the resulting soluble oligomeric THAP1 is recovered from the culture supernatant.

In some embodiments of the invention, a THAP1 or a THAP-family member dimer is created by fusing THAP1 or a THAP-family member to an Fc region polypeptide derived from an antibody, in a manner that does not substantially affect the binding of THAP1 or a THAP-family member to a chemokine, such as SLC/CCL21. Preparation of fusion proteins comprising heterologous polypeptides fused to various portions of antibody-derived polypeptides (including Fc region) has been described, e.g., by Ashkenazi et al. (1991) *PNAS* 88:10535, Byrn et al. (1990) *Nature* 344:667, and Hollenbaugh and Aruffo "Construction of Immunoglobulin Fusion Proteins", in *Current Protocols in Immunology*, Supp. 4, pages 10.19.1 - 10.19.11, 1992. The THAP-family/Fc fusion proteins are allowed to assemble much like antibody molecules, whereupon interchain disulfide bonds form between Fc polypeptides, yielding divalent THAP. Similar fusion proteins of TNF receptors and Fc (see for example Moreland et al. (1997) *N. Engl. J. Med.* 337(3):141-147; van der Poll et al. (1997) *Blood* 89(10):3727-3734; and Ammann et al. (1997) *J. Clin. Invest.* 99(7):1699-1703) have been used successfully for treating rheumatoid arthritis. Soluble derivatives have also been made of cell surface glycoproteins in the immunoglobulin gene superfamily consisting of an extracellular domain of the cell surface glycoprotein fused to an immunoglobulin constant (Fc) region (see e.g., Capon, D. J. et al. (1989) *Nature* 337:525-531 and

Capon U.S. Patent Nos. 5,116,964 and 5,428,130 [CD4-IgG1 constructs]; Linsley, P. S. et al. (1991) J. Exp. Med. 173:721-730 [a CD28-IgG1 construct and a B7-1-IgG1 construct]; and Linsley, P. S. et al. (1991) J. Exp. Med. 174:561-569 and U.S. Patent No. 5,434,131 [a CTLA4-IgG1]). Such fusion proteins have proven useful for modulating receptor-ligand interactions.

5           Some embodiments relate to THAP-immunoglobulin fusion proteins and THAP chemokine-binding domain fusions with immunoglobulin molecules or fragments thereof. Such fusions can be produced using standard methods, for example, by creating an expression vector encoding the SLC/CCL21 chemokine-binding protein THAP1 fused to the antibody polypeptide and inserting the vector into a suitable host cell. One suitable Fc polypeptide is the native Fc region  
10 polypeptide derived from a human IgG1, which is described in International Publication WO 93/10151. Another useful Fc polypeptide is the Fc mutein described in U.S. Patent No. 5,457,035. The amino acid sequence of the mutein is identical to that of the native Fc sequence presented in International Publication WO 93/10151, except that amino acid 19 has been changed from Leu to Ala, amino acid 20 has been changed from Leu to Glu, and amino acid 22 has been changed from  
15 Gly to Ala. This mutein Fc exhibits reduced affinity for immunoglobulin receptors.

          SLC/chemokine-binding fragments of human THAP1 or THAP-family polypeptides, rather than the full protein, can also be employed in methods of the invention. Fragments may be less immunogenic than the corresponding full-length proteins. The ability of a fragment to bind chemokines, such as SLC, can be determined using a standard assay. Fragments can be prepared by  
20 any of a number of conventional methods. For example, a desired DNA sequence can be synthesized chemically or produced by restriction endonuclease digestion of a full length cloned DNA sequence and isolated by electrophoresis on agarose gels. Linkers containing restriction endonuclease cleavage sites can be employed to insert the desired DNA fragment into an expression vector, or the fragment can be digested at naturally-present cleavage sites. The polymerase chain  
25 reaction (PCR) can also be employed to isolate a DNA sequence encoding a desired protein fragment. Oligonucleotides that define the termini of the desired fragment are used as 5' and 3' primers in the PCR procedure. Additionally, known mutagenesis techniques can be used to insert a stop codon at a desired point, e.g., immediately downstream of the codon for the last amino acid of the desired fragment.

30           In other embodiments, a THAP-family polypeptide or a biologically active fragment thereof, for example, an SLC-binding domain of THAP1 may be substituted for the variable portion of an antibody heavy or light chain. If fusion proteins are made with both heavy and light chains of an antibody, it is possible to form a THAP-family polypeptide oligomer with at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, or more than  
35 nine THAP-family polypeptides.

          In some embodiments of the present invention, THAP-chemokine binding can be provided to decrease the biological availability of a chemokine or otherwise disrupt the activity of

chemokine. For example, THAP-family polypeptides, SLC-binding domains of THAP-family polypeptides, THAP oligomers, and SLC-binding domain-THAP1-immunoglobulin fusion proteins of the invention can be used to interact with SLC thereby preventing it from performing its normal biological role. In some embodiments, the entire THAP1 polypeptide (SEQ ID NO: 3) can be used  
5 to bind to SLC. In other embodiments, fragments of THAP1, such as the SLC-binding domain of the THAP1 (amino acids 143-213 of SEQ ID NO: 3) can be used to bind to SLC. Such fragments can be from at least 8, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 160, at least  
10 170, at least 180, at least 190, at least 200, at least 210 or at least 213 consecutive amino acids of SEQ ID NO: 3. In some embodiments, fragments can be from at least 8, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65 or at least 70 consecutive amino acids of (amino acids 143-213 of SEQ ID NO: 3). THAP-family polypeptides that may be capable of binding SLC, for example THAP2-11  
15 and THAP0 or biologically active fragments thereof can also be used to bind to SLC so as to decrease its biological availability or otherwise disrupt the activity of this chemokine.

In some embodiments, a plurality of THAP-family proteins, such as a fusion of two or more THAP1 proteins or fragments thereof which comprise an SLC-binding domain (amino acids 143-213 of SEQ ID NO: 3) can be used to bind SLC. For example, oligomers comprising THAP1  
20 fragments of a size of at least 8, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65 or at least 70 consecutive amino acids of SEQ ID NO: 3 (amino acids 143-213) can be generated. Amino acid fragments which make up the THAP oligomer may be of the same or different lengths. In some embodiments, the entire THAP1 protein or biologically active portions thereof may be fused  
25 together to form an oligomer capable of binding to SLC. THAP-family polypeptides that may be capable of binding SLC, for example THAP2-11 and THAP0, the THAP-family polypeptides of SEQ ID NOs: 1-114 or biologically active fragments thereof can also be used to create oligomers which bind to SLC so as to decrease its biological availability or otherwise disrupt the activity of this chemokine.

30 According to another embodiment of the present invention, THAP-family proteins, such as THAP1 or portion of THAP1 which comprise an SLC binding domain (amino acids 143-213 of SEQ ID NO: 3), may be fused to an immunoglobulin or portion thereof. The portion may be an entire immunoglobulin, such as IgG, IgM, IgA or IgE. Additionally, portions of immunoglobulins, such as an Fc domain of the immunoglobulin, can be fused to a THAP-family polypeptide, such as  
35 THAP1, fragments thereof or oligomers thereof. Fragments of THAP1 can be, for example, at least 8, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65 or at least 70 consecutive amino acids of

SEQ ID NO: 3 (amino acids 143-213). In some embodiments, THAP-family polypeptides that may be capable of binding SLC, for example THAP2-11 and THAP0, the THAP-family polypeptides of SEQ ID NOs: 1-114 or biologically active fragments thereof can also be used to form immunoglobulin fusion that bind to SLC so as to decrease its biological availability or otherwise  
 5 disrupt the activity of this chemokine.

Some aspects of the present invention relate to THAP-family polypeptides, chemokine-binding domains of THAP-family polypeptides, THAP oligomers, and chemokine-binding domain-THAP-immunoglobulin fusion proteins such as those described above which bind to chemokines other than SLC. For example, THAP-family polypeptides, chemokine-binding domains of THAP-  
 10 family polypeptides, THAP oligomers, and chemokine-binding domain-THAP-immunoglobulin fusion proteins can be used to bind to or otherwise interact with chemokines from many families such as C chemokines, CC chemokines, C-X-C chemokines, C-X3-C chemokines, XC chemokines or CCK chemokines. In particular, THAP-family polypeptides, chemokine-binding domains of THAP-family polypeptides, THAP oligomers, and chemokine-binding domain-THAP-  
 15 immunoglobulin fusion proteins may interact with chemokines such as XCL1, XCL2, CCL1, CCL2, CCL3, CCL3L1, SCYA3L2, CCL4, CCL4L, CCL5, CCL6, CCL7, CCL8, SCYA9, SCYA10, CCL11, SCYA12, CCL13, CCL14, CCL15, CCL16, CCL17, CCL18, CCL19, CCL20, CCL21, CCL22, CCL23, CCL24, CCL25, CCL26, CCL27, CCL28, clone 391, CARP CC-1, CCL1, CK-1, regakine-1, K203, CXCL1, CXCL1P, CXCL2, CXCL3, PF4, PF4V1, CXCL5,  
 20 CXCL6, PPBP, SPBPBP, IL8, CXCL9, CXCL10, CXCL11, CXCL12, CXCL14, CXCL15, CXCL16, NAP-4, LFCA-1, Scyba, JSC, VHSV-induced protein, CX3CL1 and fCL1.

In some embodiments of the present invention, THAP-family polypeptides, chemokine-binding domains of THAP-family polypeptides, THAP oligomers, and chemokine-binding domain-THAP-immunoglobulin fusion proteins can bind to a chemokine extracellularly. For example, the  
 25 THAP1 polypeptide, a biologically active fragment thereof (such as the SLC-binding domain of THAP1 (amino acids 143-213 of SEQ ID NO: 3)), an oligomer thereof, or an immunoglobulin fusion thereof can bind to a chemokine extracellularly. In other examples, chemokine-binding domains of other THAP-family members such as THAP2, THAP3, THAP4, THAP5, THAP6, THAP7, THAP8, THAP9, THAP10, THAP11 or THAP0, biologically active fragments thereof,  
 30 oligomers thereof, or immunoglobulin fusions thereof can be used to bind to chemokines extracellularly. Binding of the THAP-family polypeptides, chemokine-binding domains of THAP-family polypeptides, THAP oligomers, and chemokine-binding domain-THAP-immunoglobulin fusion proteins may either decrease or increase the affinity of the chemokine for its extracellular receptor. In cases where binding of the chemokine to its extracellular receptor is inhibited, the  
 35 normal biological effect of the chemokine can be inhibited. Such inhibition can prevent the occurrence of chemokine-mediated cellular responses, such as the modulation of cell proliferation, the modulation of angiogenesis, the modulation of an inflammation response, the modulation of

apoptosis, the modulation of cell differentiation. In some embodiments, inhibition of the binding of a chemokine to its extracellular receptor can result in transcriptional modulation. Alternatively, in cases where binding of the chemokine to its extracellular receptor is activated, the normal biological effect of the chemokine can be enhanced. Such enhancement can increase the occurrence of chemokine-mediated cellular responses, such as the modulation of cell proliferation, the modulation of angiogenesis, the modulation of an inflammation response, the modulation of apoptosis, the modulation of cell differentiation. In some embodiments, enhancement of the binding of a chemokine to its extracellular receptor can result in transcriptional modulation.

In some embodiments of the present invention, THAP-family polypeptides, chemokine-binding domains of THAP-family polypeptides, THAP oligomers, and chemokine-binding domain-THAP-immunoglobulin fusion proteins can bind to a chemokine intracellularly. In some embodiments, the THAP-family protein acts as a nuclear receptor for the chemokine. For example, the THAP1 polypeptide, a biologically active fragment thereof (such as the SLC-binding domain of THAP1 (amino acids 143-213 of SEQ ID NO: 3)), an oligomer thereof, or an immunoglobulin fusion thereof can bind to a chemokine intracellularly. In other examples, chemokine-binding domains of other THAP-family members such as THAP2, THAP3, THAP4, THAP5, THAP6, THAP7, THAP8, THAP9, THAP10, THAP11 or THAP0, biologically active fragments thereof, oligomers thereof, or immunoglobulin fusions thereof can be used to bind to chemokines intracellularly. Binding of the THAP-family polypeptides, chemokine-binding domains of THAP-family polypeptides, THAP oligomers, and chemokine-binding domain-THAP-immunoglobulin fusion proteins may either decrease or increase the affinity of the chemokine for its intracellular receptor. In other embodiments, the THAP-family polypeptides, chemokine-binding domains of THAP-family polypeptides, THAP oligomers, and chemokine-binding domain-THAP-immunoglobulin fusion proteins are the intracellular receptor for the chemokine. In cases where binding of the chemokine to its intracellular receptor is inhibited, the normal biological effect of the chemokine can be inhibited. Such inhibition can prevent the occurrence of chemokine-mediated cellular responses, such as the modulation of cell proliferation, the modulation of angiogenesis, the modulation of an inflammation response, the modulation of apoptosis, the modulation of cell differentiation. In some embodiments, inhibition of the binding of a chemokine to its intracellular receptor can result in transcriptional modulation. Alternatively, in cases where binding of the chemokine to its intracellular receptor is activated, the normal biological effect of the chemokine can be enhanced. Such enhancement can increase the occurrence of chemokine-mediated cellular responses, such as the modulation of cell proliferation, the modulation of angiogenesis, the modulation of an inflammation response, the modulation of apoptosis, the modulation of cell differentiation. In some embodiments, enhancement of the binding of a chemokine to its intracellular receptor can result in transcriptional modulation.



In accordance with another aspect of the invention, THAP-family polypeptides, chemokine-binding domains of THAP-family polypeptides, THAP oligomers, and chemokine-binding domain-THAP-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions. Such pharmaceutical compositions can be used to decrease or increase the bioavailability and functionality of a chemokine. For example, THAP-family polypeptides, SLC-binding domains of THAP-family polypeptides, THAP oligomers, and SLC-binding domain-THAP1-immunoglobulin fusion proteins of the present invention can be administered to a subject to inhibit an interaction between SLC and its receptor, such as CCR7, on the surface of cells, to thereby suppress SLC-mediated responses. The inhibition of chemokine SLC may be useful therapeutically for both the treatment of inflammatory or proliferative disorders, as well as modulating (e.g., promoting or inhibiting) cell differentiation, cell proliferation, and/or cell death.

In an additional embodiment of the present invention, the THAP-family polypeptides, chemokine-binding domains of THAP-family polypeptides, THAP oligomers, and chemokine-binding domain-THAP-immunoglobulin fusion proteins of the present invention can be used to detect the presence of a chemokine in a biological sample and in screening assays to identify molecules which inhibit the interaction of a THAP-family polypeptide with a chemokine. For example, the THAP-family polypeptides, SLC-binding domains of THAP-family polypeptides, THAP oligomers, and SLC-binding domain-THAP1-immunoglobulin fusion proteins of the present invention can be used to detect the presence of SLC in a biological sample and in screening assays to identify molecules which inhibit the interaction of a THAP-family polypeptide with SLC. Such screening assays are similar to those described below for PAR4-THAP interactions.

Certain aspects of the present invention related to a method of identifying a test compound that modulates THAP-mediated activities. In some cases the THAP-mediated activity is SLC-binding. Test compounds which affect THAP-SLC binding can be identified using a screening method wherein a THAP-family polypeptide or a biologically active fragment thereof is contacted with a test compound. In some embodiments, the THAP-family polypeptide comprises an amino acid sequence having at least 30% amino acid identity to an amino acid sequence of SEQ ID NO: 1 or SEQ ID NO: 2. Whether the test compound modulates the binding of SLC with a THAP-family polypeptide, such as THAP1 (SEQ ID NO: 3), is determined by determining whether the test compound modulates the activity of the THAP-family polypeptide or biologically active fragment thereof. Biologically active fragments of a THAP-family polypeptide may be at least 5, at least 8, at least 10, at least 12, at least 15, at least 18, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 160, at least 170, at least 180, at least 190, at least 200, at least 210, at least 220 or at least more than 220 amino acids in length. A determination that the test compound modulates the activity of said polypeptide indicates that the test compound is a candidate modulator of THAP-mediated activities.

Although THAP-family polypeptides, chemokine-binding domains of THAP-family polypeptides, THAP oligomers, and chemokine-binding domain-THAP-immunoglobulin fusion proteins can be used for the above-mentioned chemokine interactions, it will be appreciated that homologs of THAP-family polypeptides, chemokine-binding domains of THAP-family polypeptides, THAP oligomers, and chemokine-binding domain-THAP-immunoglobulin fusion proteins can be used in place of THAP-family polypeptides, chemokine-binding domains of THAP-family polypeptides, THAP oligomers, and chemokine-binding domain-THAP-immunoglobulin fusion proteins. For example, homologs having at least about 30-40% identity, preferably at least about 40-50% identity, more preferably at least about 50-60%, and even more preferably at least about 60-70%, 70-80%, 80%, 90%, 95%, 97%, 98%, 99% or 99.8% identity across the amino acid sequences of SEQ ID NOs: 1-114 or portions thereof can be used.

Although this section, entitled "Oligomeric Forms of THAP-1," primarily describes THAP-family polypeptides, SLC-binding domains of THAP-family polypeptides, THAP oligomers, SLC-binding domain-THAP-immunoglobulin fusion proteins and homologs of these polypeptides as well as methods of using such polypeptides, it will be appreciated that such polypeptides are included in the class of THAP-type chemokine-binding agents. Accordingly, the above description also applies to THAP-type chemokine-binding agents. It will be appreciated that THAP-type chemokine-binding agents will be used for applications which include, but are not limited to, chemokine binding, inhibiting or enhancing chemokine activity, chemokine detection, reducing the symptoms associated with a chemokine influenced or mediated condition, and reducing or preventing inflammation or other chemokine mediated conditions. THAP-type chemokine-binding agents can also be used in the kits, devices, compositions, and procedures described elsewhere herein.

In some embodiments of the present invention, THAP-type chemokine-binding agents bind to or otherwise modulate the activity of one or more chemokines selected from the group consisting of XCL1, XCL2, CCL1, CCL2, CCL3, CCL3L1, SCYA3L2, CCL4, CCL4L, CCL5, CCL6, CCL7, CCL8, SCYA9, SCYA10, CCL11, SCYA12, CCL13, CCL14, CCL15, CCL16, CCL17, CCL18, CCL19, CCL20, CCL21, CCL22, CCL23, CCL24, CCL25, CCL26, CCL27, CCL28, clone 391, CARP CC-1, CCL1, CK-1, regakine-1, K203, CXCL1, CXCL1P, CXCL2, CXCL3, PF4, PF4V1, CXCL5, CXCL6, PPBP, SPBPBP, IL8, CXCL9, CXCL10, CXCL11, CXCL12, CXCL14, CXCL15, CXCL16, NAP-4, LFCA-1, Scyba, JSC, VHSV-induced protein, CX3CL1, and fCL1.

#### **Chemokine Binding Domains**

In some embodiments of the present invention a chemokine-binding domain that consists essentially of the chemokine binding portion of a THAP-family polypeptide is contemplated. In some embodiments, the THAP-family polypeptide is THAP-1 (SEQ ID NO: 3) or a homolog thereof. Chemokines that are capable of binding to any particular THAP-family member can be determined as described in Examples 16, 32 and 33, which set out both *in vitro* and *in vivo* assays

for determining the binding affinity of several different chemokines to THAP-1. The portion of the THAP-family protein that binds to the chemokine can readily be determined through the analysis of deletion and point mutants of any of the THAP-family members capable of chemokine-binding. Such analyses of deletion and point mutants were used to determine the specific region of THAP-1 that permits SLC-binding (see Example 15). Additionally, deletion and point mutation studies were used to determine portions of THAP-family proteins as well as specific amino acid residues that interact with PAR-4 (Examples 4-7 and 13). It will be appreciated that the methods described in these Examples can be used to precisely identify the chemokine-binding portion of any THAP-family member using any chemokine.

By "chemokine-binding domain" or "portion that binds to a chemokine" is meant a fragment which comprises 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 160, 170, 180, 190, 200, 210 or greater than 210 consecutive amino acids of a THAP-family polypeptide but less than the total number of amino acids present in the THAP-family polypeptide. In some embodiments, the THAP-family polypeptide is THAP-1 (SEQ ID NO: 3).

The complete amino acid sequence of each human THAP-family polypeptide is described in the Sequence Listing. In particular, THAP-1 is (SEQ ID NO: 3), THAP-2 is (SEQ ID NO: 4), THAP-3 is (SEQ ID NO: 5), THAP-4 is (SEQ ID NO: 6), THAP-5 is (SEQ ID NO: 7), THAP-6 is (SEQ ID NO: 8), THAP-7 is (SEQ ID NO: 9), THAP-8 is (SEQ ID NO: 10), THAP-9 is (SEQ ID NO: 11), THAP-10 is (SEQ ID NO: 12), THAP-11 is (SEQ ID NO: 13), THAP-0 is (SEQ ID NO: 14). The complete amino acid sequence of additional THAP-family polypeptides from other species are also listed in the Sequence Listing as SEQ ID NOs: 16-98. As such, the chemokine-binding portion of any of these THAP-family polypeptide sequences that are listed in the Sequence Listing is explicitly described. In particular, in some embodiments, the chemokine-binding domain is a fragment of a THAP-family chemokine-binding agent described by the formula:

for each THAP-family polypeptide, N = the number of amino acids in the full-length polypeptide; B = a number between 1 and N - 1; and E = a number between 1 and N.

For any THAP-family polypeptide, a chemokine-binding domain is specified by any consecutive sequence of amino acids beginning at an amino acid position B and ending at amino acid position E, wherein E > B.

### Methods Of Complex Formation Between A Chemokine And A THAP-Type Chemokine-Binding Agent

Some aspects of the present invention relate to methods for forming a complex between a chemokine and a THAP-type chemokine-binding agent. These methods include the step of  
5 contacting one or more chemokines with one or more THAP-type chemokine-binding agents described herein such that a complex comprising one or more chemokines and one or more THAP-type chemokine-binding agents is formed. In some embodiments, a plurality of different chemokines are contacted with one or a plurality of different THAP-type chemokine-binding agents so as to form one or more complexes. Alternatively, a plurality of different THAP-type chemokine-binding agents are contacted with one or more chemokines so as to form one or more complexes.  
10

A number of different chemokines can be used in the above-described complex formation methods. Such chemokines include, but are not limited to, XCL1, XCL2, CCL1, CCL2, CCL3, CCL3L1, SCYA3L2, CCL4, CCL4L, CCL5, CCL6, CCL7, CCL8, SCYA9, SCYA10, CCL11, SCYA12, CCL13, CCL14, CCL15, CCL16, CCL17, CCL18, CCL19, CCL20, CCL21, CCL22,  
15 CCL23, CCL24, CCL25, CCL26, CCL27, CCL28, clone 391, CARP CC-1, CCL1, CK-1, regakine-1, K203, CXCL1, CXCL1P, CXCL2, CXCL3, PF4, PF4V1, CXCL5, CXCL6, PPBP, SPBPBP, IL8, CXCL9, CXCL10, CXCL11, CXCL12, CXCL14, CXCL15, CXCL16, NAP-4, LFCA-1, Scyba, JSC, VHSV-induced protein, CX3CL1 and fCL1.

Method of forming a complex between a THAP-type chemokine-binding agent and a  
20 chemokine can be used both *in vitro* and *in vivo*. For example, *in vitro* uses can include the detection of a chemokine in a solution or a biological sample that has been removed or withdrawn from a subject. Such samples may include, but are not limited to, tissue samples, blood samples, and other fluid or solid samples of biological material. *In vivo* uses can include, but are not limited to, the detection or localization of chemokines in a subject, reducing or inhibiting the activity of one  
25 or more chemokines throughout or in certain areas of a subject's body, and reducing the symptoms associated with a chemokine influenced or mediated condition.

### Modulation of Transcription

In some embodiments of the present invention THAP-family polypeptides, THAP DNA-binding domains (THAP domains), homologs of THAP-family proteins or homologs of THAP  
30 domains are used to modulate transcription. In other embodiments, THAP-family polypeptides, THAP domains, homologs of THAP-family proteins or homologs of THAP domains interact with a chemokine to modulate transcription. In either of the above-mentioned embodiments, a THAP-family polypeptide, THAP domain, THAP-chemokine complex or homologs thereof recognize a THAP responsive element. Recognition of the THAP responsive element by a THAP-family  
35 polypeptide, THAP domain, THAP-chemokine complex or homologs thereof results in the modulation of one or more THAP responsive promoters.

As used herein, "THAP responsive promoter" means, a promoter comprising one or more THAP responsive elements. THAP responsive promoters also include promoters that are indirectly regulated by THAP. For example, a THAP responsive element may be present as an upstream enhancer sequence, the presence of which, activates transcription at the downstream promoter. In another nonlimiting example, a first promoter may be modulated by a polypeptide that is encoded by a gene under the control of a second promoter having a THAP responsive element, however, the first promoter does not comprise a THAP responsive element. In such a case, the activity of the first promoter is indirectly responsive to THAP because transcription is modulated by the polypeptide encoded by the second promoter which is responsive to THAP.

As used herein, "THAP responsive elements" include, but are not limited to, nucleic acids which comprise one or more of the following nucleotide consensus sequences. The first THAP responsive element consensus sequence comprises the nucleotide sequences GGGCAA or TGGCAA organized as direct repeats with approximately a 5 nucleotide spacing (DR-5 motifs). For example, one consensus sequence is GGGCAAnnnnnTGGCAA (SEQ ID NO: 149). Although GGGCAA and TGGCAA sequences constitute a typical THAP domain DNA binding site (THAP responsive element), GGGCAT, GGGCAG and TGGCAG sequences are also DNA target sequences recognized by the THAP DNA-binding domain. Additionally, a second THAP responsive element consensus sequence comprises the nucleotide sequences TTGCCA or GGGCAA organized as everted repeats with 11 nucleotide spacing (ER-11 motifs). For example, one consensus sequence is TTGCCAnnnnnnnnnnnTGGCAA (SEQ ID NO: 159). Although TTGCCA and GGGCAA sequences constitute a typical THAP responsive element, CTGCCA is also recognized.

Another THAP responsive element is the THRE consensus sequence which is illustrated in Figure 24 (SEQ ID NO: 306). In some embodiments of the present invention, THRE is a preferential recognition motif for monomeric THAP-family polypeptides or biologically active fragments thereof. In some embodiments, THRE is preferentially recognized by the THAP1 monomer. Alternatively, in some embodiments, the DR-5 and/or the ER-11 motif is preferentially recognized by a dimer or a multimer of a THAP-family polypeptide or biologically active fragments thereof. In some embodiments, the THAP dimers or multimers comprise THAP1.

A THAP responsive element can comprise either a single type of consensus nucleotide sequence, multiple types of consensus sequences. For example, a THAP responsive element can comprise one, two, three, four, five or more than five DR-5 consensus sequences. Similarly, a THAP responsive element can comprise one, two, three, four, five or more than five ER-11 consensus sequences. In another example, a THAP responsive element can comprise one, two, three, four, five or more than five THRE consensus sequences. In addition, a THAP responsive element can comprise a mixture of two, three, four, five or more than five DR-5, ER-11 and THRE consensus sequences. Furthermore, any of the aforementioned THAP responsive elements can

comprise one or more variants of DR-5, ER-11 or THRE consensus sequences or variants of some or all of DR-5, ER-11 or THRE consensus sequences.

It will be appreciated that other minor nucleotide sequence variations can occur in THAP responsive element consensus sequences which do not substantially affect the binding of the THAP domain to the THAP responsive element. For example, a THAP responsive element can comprise a nucleic acid having at least 99%, at least 98%, at least 97%, at least 96%, at least 95, at least 94%, at least 93%, at least 92%, at least 91%, at least 90, at least 89%, at least 88%, at least 87%, at least 86%, at least 85, at least 84%, at least 83%, at least 82%, at least 81%, at least 80, at least 75%, at least 70%, at least 65%, at least 60%, at least 55%, or at least 50% nucleotide sequence identity to a consensus sequence for DR-5, ER-11 or THRE.

In some embodiments of the present invention, the THAP-family polypeptide, THAP domain, THAP-chemokine complex or homologs thereof recognize a THAP responsive element in the promoter of the gene or genes whose transcription is modulated. Alternatively, in other embodiments, the THAP-family polypeptide, THAP domain, THAP-chemokine complex or homologs thereof recognize a THAP responsive element at locations other than the promoter of the gene or genes whose transcription is modulated.

Upon binding of the THAP responsive element by a THAP-family polypeptide, THAP domain, THAP-chemokine complex or homolog thereof transcription can be modulated. Such modulation may include repression or activation of transcription. Whether transcription is repressed or activated, as well as the extent of repression or activation, can be influenced by many factors, including but not limited to, the number and position of THAP responsive elements, the THAP-family member or homolog that is bound and, in the case of THAP-chemokine complexes, the type of chemokine that forms the THAP chemokine complex.

In some embodiments, chemokine analogs can be used to bind to THAP-family polypeptides or biologically active fragments thereof. For example, a chemokine can be modified so as to retain its THAP-binding or THAP interaction activity but alter other of its physiological effects. Such chemokine analogs can be used to modulate transcription by allowing recognition and binding of THAP to a THAP responsive element without mediating other of its physiological effects. As used herein, "chemokine analogs" are chemokine homologs having at least 99%, at least 97%, at least 95, at least 93%, at least 90, at least 85, at least 80, at least 75%, at least 70%, at least 65%, at least 60%, at least 50%, at least 40% or at least 30% amino acid identity to a specific chemokine. For example, analogs of SLC comprise polypeptide homologs of SLC having at least 99%, at least 97%, at least 95, at least 93%, at least 90, at least 85, at least 80, at least 75%, at least 70%, at least 65%, at least 60%, at least 50%, at least 40% or at least 30% amino acid identity to SLC. As another example, analogs of CXCL9 comprise polypeptide homologs of CXCL9 having at least 99%, at least 97%, at least 95, at least 93%, at least 90, at least 85, at least 80, at least 75%,

at least 70%, at least 65%, at least 60%, at least 50%, at least 40% or at least 30% amino acid identity to CXCL9. Chemokine analogs can also include chemically modified chemokines.

Some embodiments of the present invention relate to the screening of a test compound to determine whether it is capable of modulating transcription of a nucleic acid under control of a THAP responsive element. A number of constructs can be generated wherein a nucleic acid is placed under control of at least one THAP responsive element. In some embodiments, the construct is introduced into a cell which is responsive to a chemokine. For example, in some embodiments, the construct is introduced into a cell which is responsive to SLC, such as a cell expressing the CCR7 receptor. In another example, in some embodiments, the construct is introduced into a cell which is responsive to CXCL9, such as a cell expressing the CXCR3 receptor. For example, a nucleic acid can be operably linked to a promoter comprising one or more THAP responsive elements. The nucleic acid can be nucleic acid which results in a transcript that is capable of detection. The transcript may be detected and quantified by any method known in the art. In some embodiments, the nucleic acid will encode a reporter enzyme, including but not limited to, GFP, luciferase,  $\beta$ -galactosidase, and gus. The activity of such a reporter enzyme can be used to measure the amount of transcription that occurs from the promoter containing the THAP responsive elements.

In some embodiments, a THAP-family protein is allowed to contact the construct comprising the nucleic acid that is under control of the THAP responsive element. The THAP-family protein may modulate transcription in the absence of the test compound. Alternatively, the THAP-family protein may only modulate transcription in the presence of a test compound. In either case, the effect of the test compound on the modulation of transcription can be determined by determining the increase or decrease in transcription that is caused by the test compound when compared to the base level of transcription that occurs in the presence of THAP-family protein prior to the addition of test compound. Determining whether the presence of test compound increases or decrease the level of transcription at the THAP responsive element when compared to the level of transcription in the absence of test compound permits the determination of whether the compound modulates transcription of a nucleic acid under the control of a THAP responsive element.

Certain aspects of the present invention also relate to the use of THAP-family polypeptide-chemokine transcription modulators in the treatment or amelioration of conditions resulting from too much or a deficiency in the transcription of certain genes. Modulation of the interaction of a chemokine with a THAP-family polypeptide can be used in the treatment of an individual suffering from one or more specific conditions. For example, the interaction between chemokines and THAP-family members, such as the polypeptides of SEQ ID NOs: 1-114 can be used modulate transcription of certain genes thereby resulting in suppression of tumorigenesis and/or metastasis, inhibition or stimulation of apoptosis of endothelial cells in angiogenesis-dependent diseases including but not limited to cancer, cardiovascular diseases, inflammatory diseases, and inhibition

of apoptosis of neurons in acute and chronic neurodegenerative disorders, including but not limited to Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral sclerosis, HIV encephalitis, stroke, epileptic seizures and malignant tumors.

5 In some embodiments chemokine analogs can be used to interact with THAP-family polypeptides so as to treat or otherwise ameliorate the symptoms associated with the above-mentioned conditions.

It will be appreciated that THAP-type chemokine-binding agents can also be used to modulate transcription as described above. Some embodiments of such modulation of transcription are set out below.

#### 10 **Transcription Factor Decoys**

Some embodiments of the present invention relate to transcription factor decoys and methods of their use. In some embodiments of the present invention, a transcription factor decoy is any molecule that functions to inhibit or otherwise modulate the effect of a THAP/chemokine complex or a THAP-family polypeptide or a biologically active fragment thereof on gene transcription. In some embodiments, a transcription factor decoy is a molecule that acts as an inhibitor of the interaction between a THAP-family polypeptide or a biologically active fragment thereof and a nucleic acid. Alternatively, a transcription factor decoy can inhibit the interaction between a THAP/chemokine complex and a nucleic acid. For example, the nucleic acid can be a THAP responsive promoter or any other nucleic acid sequence which is involved in the modulation of the expression of a THAP responsive gene or a gene responsive to a THAP/chemokine complex.

20 In some embodiments of the present invention, the transcription factor decoy functions to inhibit, lessen or negate the effect of a THAP/chemokine complex or a THAP-family polypeptide or a biologically active fragment thereof on the expression of certain genes. For example, some transcription factor decoys function as competitive inhibitors of the interaction between a nucleic acid and a THAP/chemokine complex or a nucleic acid and a THAP-family polypeptide or a biologically active fragment thereof. In other embodiments, the transcription factor decoy functions as a nonreversible or suicide inhibitor. In yet other embodiments, the transcription factor decoy acts as a reversible inhibitor.

Some embodiments of the present invention contemplate transcription factor decoys which comprise one or more nucleic acids which comprise or consist essentially of a THAP responsive element. THAP responsive elements that are useful for the construction of transcription factor decoys include, but are not necessarily limited to, DR-5 elements, ER-11 elements and THRE elements. In some embodiments, the transcription factor decoys comprise one or more nucleic acids having a nucleotide sequence selected from the group consisting of SEQ ID NOs: 140-159 and 306. In some embodiments of the present invention, transcription factor decoys comprise a plurality of nucleic acids which comprise one or more THAP responsive elements. In such embodiments, the sequence of the THAP responsive elements may be the same or different.



Some embodiments of the present invention also contemplates pharmaceutical compositions which one or more transcription factor decoys in a pharmaceutically acceptable carrier. As described above, the pharmaceutical compositions can comprise transcription factor decoys comprising one or more nucleic acid sequences which comprise one or more THAP responsive elements.

Additional embodiments of the present invention contemplate methods of using transcription factor decoys to inhibit, lessen or otherwise modulate the expression of one or more genes that are responsive to a THAP/chemokine complex or one or more genes that are responsive to a THAP-family polypeptide or a fragment thereof.

#### 10 **Effect of Interactions Between Chemokines and Thap-Type Chemokine-Binding Agents**

Some embodiments of the present invention relate to methods of modulating chemokine interactions with cellular receptors. Such receptors can be extracellular or can be molecules that are present within the cell. For example, chemokines SLC and ELC can bind to extracellular chemokine receptors CCR7 and CCR11. The chemokine CCL5 binds to extracellular chemokine receptors CCR1, CCR3 and CCR5. The CXCL-family chemokines, CXCL9 and CXCL10, bind to the extracellular chemokine receptor, CXCR3. Other chemokine interactions with receptors are also known in the art and are included in Ransohoff, R. M. and Karpus, W. J. (2001). Roles of Chemokines and Their Receptors in the Induction and Regulation of Autoimmune Disease, in Contemporary Clinical Neuroscience: Cytokines and Autoimmune Diseases, V. K. Kuchroo, et al., eds. Humana Press, Totowa, N.J., pages 157-191.

In some embodiments of the present invention the interaction of chemokines with extracellular receptors are enhanced or inhibited by providing to a cell, which expresses one or more extracellular chemokine receptors, a THAP-type chemokine-binding agent. Such extracellular receptors can include, but are not limited to, CCR1, CCR2, CCR3, CCR4, CCR5, CCR6, CCR7, CCR8, CCR9, CCR10, CCR11, CXCR1, CXCR2, CXCR3, CXCR4 and CXCR5. In some embodiments of the present invention, a THAP-type chemokine-binding agent binds to or otherwise interacts with a chemokine thereby forming a complex which binds to the extracellular receptor with more or less affinity. In some embodiments, chemokine interaction with one or more extracellular receptors is modulated by providing one or more THAP-type chemokine-binding agents.

Other aspects of the present invention relate to modulating the movement of a chemokine from the outside of a cell to the inside of the cell. For example, modulation of chemokine interaction with one or more extracellular receptors can increase or decrease the uptake of chemokines into the cell. In some embodiments of the present invention, chemokine uptake into a cell is modulated by providing THAP-type chemokine-binding agent either *in vitro* or *in vivo* in the proximity of cell which expresses one or more chemokine receptors. The THAP-type chemokine-binding agent binds to or otherwise interacts with one or more chemokines including, but not

limited to, XCL1, XCL2, CCL1, CCL2, CCL3, CCL3L1, SCYA3L2, CCL4, CCL4L, CCL5, CCL6, CCL7, CCL8, SCYA9, SCYA10, CCL11, SCYA12, CCL13, CCL14, CCL15, CCL16, CCL17, CCL18, CCL19, CCL20, CCL21, CCL22, CCL23, CCL24, CCL25, CCL26, CCL27, CCL28, clone 391, CARP CC-1, CCL1, CK-1, regakine-1, K203, CXCL1, CXCL1P, CXCL2, CXCL3, PF4, PF4V1, CXCL5, CXCL6, PPBP, SPBPBP, IL8, CXCL9, CXCL10, CXCL11, CXCL12, CXCL14, CXCL15, CXCL16, NAP-4, LFCA-1, Scyba, JSC, VHSV-induced protein, CX3CL1 and fCL1 thereby modulating the uptake of the chemokine into the cell.

In some embodiments, THAP-type chemokine-binding agents form a complex with one or more chemokines inside the cell nucleus. In such embodiments, a THAP-type chemokine-binding agent is provided to a cell such that the THAP-type chemokine-binding agent binds to or otherwise interacts with one or more chemokines. The THAP-type chemokine-binding agent can be provided to cells both *in vitro* and *in vivo*. In some embodiments, the THAP-type chemokine-binding agent is provided extracellularly wherein it is taken up by the cell either prior to or after binding to a chemokine. In other embodiments, a the THAP-type chemokine-binding agent is provided inside the cell. For example, a nucleic acid encoding a THAP-type chemokine-binding agent is introduced into a cell such that the THAP-type chemokine-binding agent is expressed inside the cell. Methods of introducing expressible recombinant nucleic acids into a cell are well known in the art. In some embodiments of the present invention, the nucleic acid encoding the THAP-type chemokine-binding agent is placed under the control of a constitutive promoter. In other embodiments, the promoter which controls expression of the THAP-type chemokine-binding agent is regulatable. Chemokines which contact or enter the nucleus are bound by THAP-type chemokine-binding agent with has been introduced into the cell. For example, a nucleic acid encoding a full-length THAP1 polypeptide can be placed under control of a regulatable promoter such that, upon induction, the polypeptide is expressed then localized to the nucleus. The THAP1 that is present in the nucleus binds to SLC which has been transported to the nucleus thereby forming a THAP1/SLC complex. It will be appreciated that other methods can also be used to introduce THAP-type chemokine-binding agents into a cell. Additionally, it will be appreciated that more than one type of THAP-type chemokine-binding agent can be introduced into a cell.

In some embodiments, THAP-type chemokine-binding agents can be introduced into the cytoplasm of the cell. In such embodiments, the THAP-type chemokine-binding agents that are present in the cytoplasm of the cell can be used in the formation of complexes with one or more chemokines. The formation of such complexes modulate the transport of chemokine into the nucleus.

In some embodiments of the present invention, chemokines or complexes comprising chemokines and THAP-type chemokine-binding agents that are present within the nucleus of the cell modulate gene expression. In such embodiments, the expression of one or more genes which are under the control of a THAP responsive promoter are modulated. In some embodiments, a

THAP responsive promoter includes one or more THAP responsive elements. In other embodiments, a THAP responsive promoter need not comprise a THAP responsive element, but rather, the promoter is responsive to a gene product that is produced by a gene that is under the control of a promoter containing one or more THAP responsive elements. Such THAP responsive promoters have been described in detail above.

The THAP-type chemokine-binding agent that is used to modulate transcription of a THAP responsive promoter can be any THAP-type chemokine-binding agent; however, some preferred agents include THAP1 and polypeptides comprising an amino acid sequence having at least 99%, at least 98%, at least 97%, at least 96%, at least 95%, at least 94%, at least 93%, at least 92%, at least 91%, at least 90%, at least 89%, at least 88%, at least 87%, at least 86%, at least 85%, at least 84%, at least 83%, at least 82%, at least 81%, at least 80%, at least 75%, at least 70%, at least 65%, at least 60%, at least 55%, at least 50%, at least 45%, at least 40%, at least 35%, or at least 30% amino acid sequence identity with the amino acid of SEQ ID NO: 3. In other embodiments, the THAP-type chemokine-binding agent is a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114 or homologs thereof.

Chemokines which are useful in the modulation of transcription can be any chemokine which binds to or otherwise interacts with a THAP-type chemokine-binding agent. Such chemokines include, but are not limited to, XCL1, XCL2, CCL1, CCL2, CCL3, CCL3L1, SCYA3L2, CCL4, CCL4L, CCL5, CCL6, CCL7, CCL8, SCYA9, SCYA10, CCL11, SCYA12, CCL13, CCL14, CCL15, CCL16, CCL17, CCL18, CCL19, CCL20, CCL21, CCL22, CCL23, CCL24, CCL25, CCL26, CCL27, CCL28, clone 391, CARP CC-1, CCL1, CK-1, regakine-1, K203, CXCL1, CXCL1P, CXCL2, CXCL3, PF4, PF4V1, CXCL5, CXCL6, PPBP, SPBPBP, IL8, CXCL9, CXCL10, CXCL11, CXCL12, CXCL14, CXCL15, CXCL16, NAP-4, LFCA-1, Scyba, JSC, VHSV-induced protein, CX3CL1 and fCL1. In some embodiments, polypeptides that are homologous to one or more of the above-described chemokines can form a complex with a THAP-type chemokine-binding agent thereby modulating transcription at a THAP responsive promoter. Such homologs can include polypeptides comprising an amino acid sequence having at least 99%, at least 98%, at least 97%, at least 96%, at least 95%, at least 94%, at least 93%, at least 92%, at least 91%, at least 90%, at least 89%, at least 88%, at least 87%, at least 86%, at least 85%, at least 84%, at least 83%, at least 82%, at least 81%, at least 80%, at least 75%, at least 70%, at least 65%, at least 60%, at least 55%, at least 50%, at least 45%, at least 40%, at least 35%, or at least 30% amino acid sequence identity with the amino acid sequence of any of the above-described chemokines. In some preferred embodiments of the present invention, one or more chemokines having an amino acid sequence selected from the group consisting of SEQ ID NOs: 271, 273, 275, 277 and 289 form a complex with one or more THAP-type chemokine-binding agents thereby modulating transcription at a THAP responsive promoter. In other embodiments, chemokines comprising an amino acid sequence having at least 99%, at least 98%, at least 97%, at least 96%, at

least 95%, at least 94%, at least 93%, at least 92%, at least 91%, at least 90%, at least 89%, at least 88%, at least 87%, at least 86%, at least 85%, at least 84%, at least 83%, at least 82%, at least 81%, at least 80%, at least 75%, at least 70%, at least 65%, at least 60%, at least 55%, at least 50%, at least 45%, at least 40%, at least 35%, or at least 30% amino acid sequence identity with the amino acid sequence of a chemokine selected from the group consisting of SEQ ID NOs: 271, 273, 275, 277 and 289 form a complex with one or more THAP-type chemokine-binding agents thereby modulating transcription at a THAP responsive promoter.

#### Primers and probes

Primers and probes of the invention can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphodiester method of Narang SA et al (Methods Enzymol 1979;68:90-98), the phosphodiester method of Brown EL et al (Methods Enzymol 1979;68:109-151), the diethylphosphoramidite method of Beaucage et al (Tetrahedron Lett 1981, 22: 1859-1862) and the solid support method described in EP 0 707 592.

Detection probes are generally nucleic acid sequences or uncharged nucleic acid analogs such as, for example peptide nucleic acids which are disclosed in International Patent Application WO 92/20702, morpholino analogs which are described in U.S. Patents Numbered 5,185,444; 5,034,506 and 5,142,047. If desired, the probe may be rendered "non-extendable" in that additional dNTPs cannot be added to the probe. In and of themselves analogs usually are non-extendable and nucleic acid probes can be rendered non-extendable by modifying the 3' end of the probe such that the hydroxyl group is no longer capable of participating in elongation. For example, the 3' end of the probe can be functionalized with the capture or detection label to thereby consume or otherwise block the hydroxyl group.

Any of the polynucleotides of the present invention can be labeled, if desired, by incorporating any label known in the art to be detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive substances (including, <sup>32</sup>P, <sup>35</sup>S, <sup>3</sup>H, <sup>125</sup>I), fluorescent dyes (including, 5-bromodesoxyuridin, fluorescein, acetylaminofluorene, digoxigenin) or biotin. Preferably, polynucleotides are labeled at their 3' and 5' ends. Examples of non-radioactive labeling of nucleic acid fragments are described in (Urdea et al. (Nucleic Acids Research. 11:4937-4957, 1988) or Sanchez-Pescador et al. (J. Clin. Microbiol. 26(10):1934-1938, 1988). In addition, the probes according to the present invention may have structural characteristics such that they allow the signal amplification, such structural characteristics being, for example, branched DNA probes as those described by Urdea et al (Nucleic Acids Symp. Ser. 24:197-200, 1991) or in the European patent No. EP 0 225 807 (Chiron).

A label can also be used to capture the primer, so as to facilitate the immobilization of either the primer or a primer extension product, such as amplified DNA, on a solid support. A capture label is attached to the primers or probes and can be a specific binding member which

forms a binding pair with the solid's phase reagent's specific binding member (e.g. biotin and streptavidin). Therefore depending upon the type of label carried by a polynucleotide or a probe, it may be employed to capture or to detect the target DNA. Further, it will be understood that the polynucleotides, primers or probes provided herein, may, themselves, serve as the capture label.

5 For example, in the case where a solid phase reagent's binding member is a nucleic acid sequence, it may be selected such that it binds a complementary portion of a primer or probe to thereby immobilize the primer or probe to the solid phase. In cases where a polynucleotide probe itself serves as the binding member, those skilled in the art will recognize that the probe will contain a sequence or "tail" that is not complementary to the target. In the case where a polynucleotide  
10 primer itself serves as the capture label, at least a portion of the primer will be free to hybridize with a nucleic acid on a solid phase. DNA labeling techniques are well known to the skilled technician.

The probes of the present invention are useful for a number of purposes. They can be notably used in Southern hybridization to genomic DNA. The probes can also be used to detect PCR amplification products. They may also be used to detect mismatches in a THAP-family gene  
15 or mRNA using other techniques.

Any of the nucleic acids, polynucleotides, primers and probes of the present invention can be conveniently immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, nitrocellulose strips, membranes, microparticles such as latex particles, sheep (or other animal) red  
20 blood cells, duracytes and others. The solid support is not critical and can be selected by one skilled in the art. Thus, latex particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and duracytes are all suitable examples. Suitable methods for immobilizing nucleic acids on solid phases include ionic, hydrophobic, covalent interactions and  
25 the like. A solid support, as used herein, refers to any material which is insoluble, or can be made insoluble by a subsequent reaction. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. Alternatively, the solid phase can retain an additional receptor which has the ability to attract and immobilize the capture reagent. The additional receptor can include a charged substance that is oppositely charged with respect to the capture reagent itself or to  
30 a charged substance conjugated to the capture reagent. As yet another alternative, the receptor molecule can be any specific binding member which is immobilized upon (attached to) the solid support and which has the ability to immobilize the capture reagent through a specific binding reaction. The receptor molecule enables the indirect binding of the capture reagent to a solid support material before the performance of the assay or during the performance of the assay. The  
35 solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cells, duracytes and other configurations known to those of ordinary

skill in the art. The nucleic acids, polynucleotides, primers and probes of the invention can be attached to or immobilized on a solid support individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the invention to a single solid support. In addition, polynucleotides other than those of the invention may be attached to the same solid support as one  
5 or more polynucleotides of the invention.

Any polynucleotide provided herein may be attached in overlapping areas or at random locations on a solid support. Alternatively the polynucleotides of the invention may be attached in an ordered array wherein each polynucleotide is attached to a distinct region of the solid support which does not overlap with the attachment site of any other polynucleotide. Preferably, such an  
10 ordered array of polynucleotides is designed to be "addressable" where the distinct locations are recorded and can be accessed as part of an assay procedure. Addressable polynucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. The knowledge of the precise location of each polynucleotides location makes these "addressable" arrays particularly useful in hybridization  
15 assays. Any addressable array technology known in the art can be employed with the polynucleotides of the invention. One particular embodiment of these polynucleotide arrays is known as the Genechips, and has been generally described in US Patent 5,143,854; PCT publications WO 90/15070 and 92/10092.

#### **Recombinant Expression Vectors and Host Cells**

20 Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof.

Vectors may have particular use in the preparation of a recombinant protein of the invention, or for use in gene therapy. Gene therapy presents a means to deliver a THAP family or  
25 THAP domain polypeptide, or a biologically active fragment or homologue thereof to a subject in order to regulate apoptosis for treatment of a disorder.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another  
30 type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome.  
35 Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In

the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

5       The recombinant expression vectors of the invention comprise a THAP-family or THAP domain nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector,  
10 "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (for example, in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory  
15 sequences are described, for example, in Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the  
20 expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., THAP-family proteins, mutant forms of THAP-family proteins, fusion proteins, or fragments of any of the preceding proteins, etc.).

25       The recombinant expression vectors of the invention can be designed for expression of a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof in prokaryotic or eukaryotic cells. For example, THAP-family or THAP domain proteins can be expressed in bacterial cells such as E. coli, insect cells (using baculovirus expression vectors) yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel, Gene Expression  
30 Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in E. coli with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-  
35 fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant

protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D. B. and Johnson, K. S. (1988) Gene 67:31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.), which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

10 Purified fusion proteins can be utilized in THAP-family activity assays, (for example, direct assays or competitive assays described in detail below), or to generate antibodies specific for THAP-family or THAP domain proteins, for example. In a preferred embodiment, a THAP-family or THAP domain fusion protein expressed in a retroviral expression vector of the present invention can be utilized to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (for example, six (6) weeks).

20 Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) Gene 69:301-315) and pET 11d (Studier et al., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 *gn10*-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 *gn1*). This viral polymerase is supplied by host strains BL21 (DE3) or HMS174(DE3) from a resident prophage harboring a T7 *gn1* gene under the transcriptional control of the lacUV 5 promoter.

25 One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al., (1992) Nucleic Acids Res. 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

35 In another embodiment, the THAP-family expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec 1 (Baldari, et al., (1987) Embo J. 6:229-234), pMFa (Kurjan and Herskowitz, (1982) Cell 30:933-943), pJRY88 (Schultz et al., (1987) Gene 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (Invitrogen Corp, San Diego, Calif.).



Alternatively, THAP-family or THAP domain proteins can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al. (1983) Mol. Cell Biol. 3:2156-2165) and the pVL series (Lucklow and Summers (1989) Virology 170:31-39). In particularly preferred embodiments, THAP-family proteins are expressed according to Karniski et al, Am. J. Physiol. (1998) 275: F79-87.

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) Nature 329:840) and pMT2PC (Kaufman et al. (1987) EMBO J. 6:187-195).

When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989. In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art, and are further described below.

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to THAP-family mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. et al., Antisense RNA as a molecular tool for genetic analysis, Reviews--Trends in Genetics, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such term refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur

in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

5 A host cell can be any prokaryotic or eukaryotic cell. For example, a THAP-family protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells or human cells). Other suitable host cells are known to those skilled in the art, including mouse 3T3 cells as further described in the Examples.

10 Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

15 For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a THAP-family protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

25 A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) a THAP-family protein. Accordingly, the invention further provides methods for producing a THAP-family protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a THAP-family protein has been introduced) in a suitable medium such that a THAP-family protein is produced. In another embodiment, the method further comprises isolating a THAP-family protein from the medium or the host cell.

30 In another embodiment, the invention encompasses a method comprising: providing a cell capable of expressing a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof, culturing said cell in a suitable medium such that a THAP-family or THAP domain protein is produced, and isolating or purifying the THAP-family or THAP domain protein from the medium or cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals, such as for the study of disorders in which THAP family proteins are implicated. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which THAP-family- or THAP domain- coding sequences have been introduced. Such host cells  
5 can then be used to create non-human transgenic animals in which exogenous THAP-family or THAP domain sequences have been introduced into their genome or homologous recombinant animals in which endogenous THAP-family or THAP domain sequences have been altered. Such animals are useful for studying the function and/or activity of a THAP-family or THAP domain polypeptide or fragment thereof and for identifying and/or evaluating modulators of a THAP-family  
10 or THAP domain activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in  
15 the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous THAP-family or THAP domain gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a  
20 cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Pat. Nos. 4,736,866 and 4,870,009, both by Leder et al., U.S. Pat. No. 4,873,191 by Wagner et al. and in Hogan, B., Manipulating the Mouse Embryo, (Cold Spring Harbor  
25 Laboratory Press, Cold Spring Harbor, N.Y., 1986).

#### Gene Therapy Vectors

Preferred vectors for administration to a subject can be constructed according to well known methods. Vectors will comprise regulatory elements (e.g. promoter, enhancer, etc) capable of directing the expression of the nucleic acid in the targeted cell. Thus, where a human cell is  
30 targeted, it is preferable to position the nucleic acid coding region adjacent to and under the control of a promoter that is capable of being expressed in a human cell.

In various embodiments, the human cytomegalovirus (CMV) immediate early gene promoter, the SV40 early promoter, the Rous sarcoma virus long terminal repeat, P actin, rat insulin promoter and glyceraldehyde-3 -phosphate dehydrogenase can be used to obtain high-level  
35 expression of the coding sequence of interest. The use of other viral or mammalian cellular or bacterial phage promoters which are well-known in the art to achieve expression of a coding sequence of interest is contemplated as well, provided that the levels of expression are sufficient for

a given purpose. By employing a promoter with well-known properties, the level and pattern of expression of the protein of interest following transfection or transformation can be optimized.

Selection of a promoter that is regulated in response to specific physiologic or synthetic signals can permit inducible expression of the gene product. For example in the case where expression of a transgene, or transgenes when a multicistronic vector is utilized, is toxic to the cells in which the vector is produced in, it may be desirable to prohibit or reduce expression of one or more of the transgenes. Several inducible promoter systems are available for production of viral vectors where the transgene product may be toxic.

The ecdysone system (Invitrogen, Carlsbad, CA) is one such system. This system is designed to allow regulated expression of a gene of interest in mammalian cells. It consists of a tightly regulated expression mechanism that allows virtually no basal level expression of the transgene, but over 200-fold inducibility. The system is based on the heterodimeric ecdysone receptor of *Drosophila*, and when ecdysone or an analog such as muristerone A binds to the receptor, the receptor activates a promoter to turn on expression of the downstream transgene high levels of mRNA transcripts are attained. In this system, both monomers of the heterodimeric receptor are constitutively expressed from one vector, whereas the ecdysone-responsive promoter which drives expression of the gene of interest is on another plasmid. Engineering of this type of system into the gene transfer vector of interest would therefore be useful. Cotransfection of plasmids containing the gene of interest and the receptor monomers in the producer cell line would then allow for the production of the gene transfer vector without expression of a potentially toxic transgene. At the appropriate time, expression of the transgene could be activated with ecdysone or muristerone A. Another inducible system that would be useful is the Tet-Off or Tet On system (Clontech, Palo Alto, CA) originally developed by Gossen and Bujard (Gossen and Bujard, 1992; Gossen et al, 1995). This system also allows high levels of gene expression to be regulated in response to tetracycline or tetracycline derivatives such as doxycycline. In the Tet-On system, gene expression is turned on in the presence of doxycycline, whereas in the Tet-Off system, gene expression is turned on in the absence of doxycycline. These systems are based on two regulatory elements derived from the tetracycline resistance operon of *E. coli*. The tetracycline operator sequence to which the tetracycline repressor binds, and the tetracycline repressor protein. The gene of interest is cloned into a plasmid behind a promoter that has tetracycline-responsive elements present in it. A second plasmid contains a regulatory element called the tetracycline-controlled transactivator, which is composed, in the Tet Off system, of the VP16 domain from the herpes simplex virus and the wild-type tetracycline repressor.

Thus in the absence of doxycycline, transcription is constitutively on. In the Tet-On system, the tetracycline repressor is not wild-type and in the presence of doxycycline activates transcription. For gene therapy vector production, the Tet Off system would be preferable so that the producer cells could be grown in the presence of tetracycline or doxycycline and prevent

expression of a potentially toxic transgene, but when the vector is introduced to the patient, the gene expression would be constitutively on.

In some circumstances, it may be desirable to regulate expression of a transgene in a gene therapy vector. For example, different viral promoters with varying strengths of activity may be utilized depending on the level of expression desired. In mammalian cells, the CMV immediate early promoter is often used to provide strong transcriptional activation. Modified versions of the CMV promoter that are less potent have also been used when reduced levels of expression of the transgene are desired. When expression of a transgene in hematopoietic cells is desired, retroviral promoters such as the LTRs from MLV or MMTV are often used. Other viral promoters that may be used depending on the desired effect include SV40, RSV LTR, HIV-1 and HIV-2 LTR, adenovirus promoters such as from the E1A, E2A, or MLP region, AAV LTR, cauliflower mosaic virus, HSV-TK, and avian sarcoma virus.

Similarly tissue specific promoters may be used to effect transcription in specific tissues or cells so as to reduce potential toxicity or undesirable effects to non-targeted tissues. For example, promoters such as the PSA, probasin, prostatic acid phosphatase or prostate-specific glandular kallikrein (hK2) may be used to target gene expression in the prostate. Similarly, promoters as follows may be used to target gene expression in other tissues.

Tissue specific promoters include in (a) *pancreas*: insulin, elastin, amylase, pdx-I, pdx-I, glucokinase; (b) *liver*: albumin PEPCK, HBV enhancer, alpha fetoprotein, apolipoprotein C, alpha-I antitrypsin, vitellogenin, NF-AB, Transthyretin; (c) *skeletal muscle*: myosin H chain, muscle creatine kinase, dystrophin, calpain p94, skeletal alpha-actin, fast troponin I; (d) *skin*: keratin K6, keratin KI; (e) *lung*: CFTR, human cytokeratin 18 (K 18), pulmonary surfactant proteins A, B and C, CC-10, Pi; (f) *smooth muscle*: sm22 alpha, SM-alpha-actin; (g) *endothelium*: endothelin- I, E-selectin, von Willebrand factor, TIE (Korhonen et al., 1995), KDR/flk-I; (h) *melanocytes*: tyrosinase; (i) *adipose tissue*: lipoprotein lipase (Zechner et al., 1988), adiponectin (Spiegelman et al., 1989), acetyl-CoA carboxylase (Pape and Kim, 1989), glycerophosphate dehydrogenase (Dani et al., 1989), adipocyte P2 (Hunt et al., 1986); and (j) *blood*: P-globin.

In certain indications, it may be desirable to activate transcription at specific times after administration of the gene therapy vector. This may be done with such promoters as those that are hormone or cytokine regulatable. For example in gene therapy applications where the indication is in a gonadal tissue where specific steroids are produced or routed to, use of androgen or estrogen regulated promoters may be advantageous. Such promoters that are hormone regulatable include MMTV, MT-1, ecdysone and RuBisco. Other hormone regulated promoters such as those responsive to thyroid, pituitary and adrenal hormones are expected to be useful in the present invention. Cytokine and inflammatory protein responsive promoters that could be used include K and T Kininogen (Kageyama et al., 1987), c-fos, TNF-alpha, C-reactive protein (Arcone et al., 1988), haptoglobin (Oliviero et al., 1987), serum amyloid A2, C/EBP alpha, IL-1, IL-6 (Poli and

Cortese, 1989), Complement C3 (Wilson et al., 1990), IL-8, alpha-1 acid glycoprotein (Prowse and Baumann, 1988), alpha-1 antitrypsin, lipoprotein lipase (Zechner et al., 1988), angiotensinogen (Ron et al., 1991), fibrinogen, c-jun (inducible by phorbol esters, TNF alpha, UV radiation, retinoic acid, and hydrogen peroxide), collagenase (induced by phorbol esters and retinoic acid), metallothionein  
5 (heavy metal and glucocorticoid inducible), Stromelysin (inducible by phorbol ester, interleukin-1 and EGF), alpha-2 macroglobulin and alpha-1 antichymotrypsin.

It is envisioned that cell cycle regulatable promoters may be useful in the present invention. For example, in a bi-cistronic gene therapy vector, use of a strong CMV promoter to drive expression of a first gene such as p16 that arrests cells in the G1 phase could be followed by  
10 expression of a second gene such as p53 under the control of a promoter that is active in the G1 phase of the cell cycle, thus providing a "second hit" that would push the cell into apoptosis. Other promoters such as those of various cyclins, PCNA, galectin-3, E2F1, p53 and BRCA1 could be used.

Tumor specific promoters such as osteocalcin, hypoxia-responsive element (HRE), NIAGE-4, CEA, alpha-fetoprotein, GRP78/BiP and tyrosinase also may be used to regulate gene  
15 expression in tumor cells. Other promoters that could be used according to the present invention include Lac-regulatable, chemotherapy inducible (e.g. MDR), and heat (hyperthermia) inducible promoters, Radiation-inducible (e.g., EGR (Joki et al., 1995)), Alpha-inhibin, RNA pol III tRNA met and other amino acid promoters, U1 snRNA (Bartlett et al., 1996), MC-1, PGK, -actin and alpha-globin. Many other promoters that may be useful are listed in Walther and Stein (1996).

20 It is envisioned that any of the above promoters alone or in combination with another may be useful according to the present invention depending on the action desired.

In addition, this list of promoters should not be considered to be exhaustive or limiting, those of skill in the art will know of other promoters that may be used in conjunction with the THAP-family and THAP domain nucleic acids and methods disclosed herein.

## 25 **Enhancers**

Enhancers are genetic elements that increase transcription from a promoter located at a distant position on the same molecule of DNA. Enhancers are organized much like promoters. That is, they are composed of many individual elements, each of which binds to one or more transcriptional proteins. The basic distinction between enhancers and promoters is operational. An  
30 enhancer region as a whole must be able to stimulate transcription at a distance; this need not be true of a promoter region or its component elements. On the other hand, a promoter must have one or more elements that direct initiation of RNA synthesis at a particular site and in a particular orientation, whereas enhancers lack these specificities. Promoters and enhancers are often overlapping and contiguous, often seeming to have a very similar modular organization.

35 Below is a list of promoters additional to the tissue specific promoters listed above, cellular promoters/enhancers and inducible promoters/enhancers that could be used in combination with the nucleic acid encoding a gene of interest in an expression construct (list of enhancers, and Table 1).

Additionally, any promoter/enhancer combination (as per the Eukaryotic Promoter Data Base EPDB) could also be used to drive expression of the gene. Eukaryotic cells can support cytoplasmic transcription from certain bacterial promoters if the appropriate bacterial polymerase is provided, either as part of the delivery complex or as an additional genetic expression construct.

- 5 Suitable enhancers include: Immunoglobulin Heavy Chain; Immunoglobulin Light Chain; T-Cell Receptor; HLA DQ (x and DQ beta; beta-Interferon; Interleukin-2; Interleukin-2 Receptor; MHC Class II 5; MHC Class II HLA-DRalpha; beta-Actin; Muscle Creatine Kinase; Prealbumin (Transthyretin); Elastase I; Metallothionein; Collagenase; Albumin Gene; alpha-Fetoprotein; -
- 10 Globin; beta-Globin; e-fos; c-HA-ras; Insulin; Neural Cell Adhesion Molecule (NCAM); alpha a1-Antitrypsin; H2B (TH2B) Histone; Mouse or Type I Collagen; Glucose-Regulated Proteins (GRP94 and GRP78); Rat Growth Hormone; Human Serum Amyloid A (SAA); Troponin I (TN I); Platelet-Derived Growth Factor; Duchenne Muscular Dystrophy; SV40; Polyoma; Retroviruses; THAPilloma Virus; Hepatitis B Virus; Human Immunodeficiency Virus; Cytomegalovirus; and Gibbon Ape Leukemia Virus.

15

TABLE 1

Element	Inducer
MT 11	Phorbol Ester (TPA)
Heavy metals MMTV (mouse mammary tumor Glucocorticoids virus)	
B-Interferon	poly(rI)X; poly(rc)
Adenovirus 5 E2	Ela
c-jun	Phorbol Ester (TPA), H2O2
H2O2 Collagenase	Phorbol Ester (TPA)
Stromelysin	Phorbol Ester (TPA), IL-1
SV40	Phorbol Ester (TPA)
Murine MX Gene	Interferon, Newcastle Disease Virus
GRP78 Gene	A23187
oc-2-Macroglobulin	IL-6
Vimentin Serum NMC Class I Gene H-2kB	Interferon
HSP70	Ela, SV40 Large T Antigen
Insulin E Box	Glucose
Proliferin	Phorbol Ester-TPA
Tumor Necrosis Factor	FMA
Thyroid Stimulating Hormone alpha Gene	Thyroid Hormone

- 20 In preferred embodiments of the invention, the expression construct comprises a virus or engineered construct derived from a viral genome. The ability of certain viruses to enter cells via receptor-mediated endocytosis and to integrate into host cell genome and express viral genes stably and efficiently have made them attractive candidates for the transfer of foreign genes into mammalian cells (Ridgeway, 1988; Nicolas and Rubenstein, 1988; Baichwal and Sugden, 1986;

Temin, 1986). The first viruses used as gene vectors were DNA viruses including the papovaviruses (simian virus 40, bovine papilloma virus, and polyoma) (Ridgeway, 1988; Baichwal and Sugden, 1986) and adenoviruses (Ridgeway, 1988; Baichwal and Sugden, 1986). These have a relatively low capacity for foreign DNA sequences and have a restricted host spectrum.

- 5 Furthermore, their oncogenic potential and cytopathic effects in permissive cells raise safety concerns. They can accommodate only up to 8 kB of foreign genetic material but can be readily introduced in a variety of cell lines and laboratory animals (Nicolas and Rubenstein, 1988; Temin, 1986).

#### *Polyadenylation Signals*

- 10 Where a cDNA insert is employed, one will typically desire to include a polyadenylation signal to effect proper polyadenylation of the gene transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and any such sequence may be employed such as human or bovine growth hormone and SV40 polyadenylation signals. Also contemplated as an element of the expression cassette is a terminator. These elements  
15 can serve to enhance message levels and to minimize read through from the cassette into other sequences.

#### *Antisense Constructs*

- The term "antisense nucleic acid" is intended to refer to the oligonucleotides complementary to the base sequences of DNA and RNA. Antisense oligonucleotides, when  
20 introduced into a target cell, specifically bind to their target nucleic acid and interfere with transcription, RNA processing, transport and/or translation. Targeting double-stranded (ds) DNA with oligonucleotide leads to triple-helix formation; targeting RNA will lead to double-helix formation.

- Antisense constructs may be designed to bind to the promoter and other control regions,  
25 exons, introns or even exon-intron boundaries of a gene. Antisense RNA constructs, or DNA encoding such antisense RNAs, may be employed to inhibit gene transcription or translation or both within a host cell, either in vitro or in vivo, such as within a host animal, including a human subject. Nucleic acid sequences comprising complementary nucleotides" are those which are capable of base-pairing according to the standard Watson-Crick complementary rules. That is, that the larger  
30 purines will base pair with the smaller pyrimidines to form only combinations of guanine paired with cytosine (G:C) and adenine paired with either thymine (A:T), in the case of DNA, or adenine paired with uracil (A:U) in the case of RNA.

- As used herein, the terms "complementary" or "antisense sequences" mean nucleic acid sequences that are substantially complementary over their entire length and have very few base  
35 mismatches. For example, nucleic acid sequences of fifteen bases in length may be termed complementary when they have a complementary nucleotide at thirteen or fourteen positions with only single or double mismatches. Naturally, nucleic acid sequences which are "completely



complementary" will be nucleic acid sequences which are entirely complementary throughout their entire length and have no base mismatches.

While all or part of the gene sequence may be employed in the context of antisense construction, statistically, any sequence 17 bases long should occur only once in the human genome and, therefore, suffice to specify a unique target sequence.

Although shorter oligomers are easier to make and increase in vivo accessibility, numerous other factors are involved in determining the specificity of hybridization. Both binding affinity and sequence specificity of an oligonucleotide to its complementary target increases with increasing length. It is contemplated that oligonucleotides of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more base pairs will be used. One can readily determine whether a given antisense nucleic acid is effective at targeting of the corresponding host cell gene simply by testing the constructs in vitro to determine whether the endogenous gene's function is affected or whether the expression of related genes having complementary sequences is affected.

In certain embodiments, one may wish to employ antisense constructs which include other elements, for example, those which include C-5 propyne pyrimidines.

Oligonucleotides which contain C-5 propyne analogues of uridine and cytidine have been shown to bind RNA with high affinity and to be potent antisense inhibitors of gene expression (Wagner et al, 1993).

#### *Ribozyme Constructs*

As an alternative to targeted antisense delivery, targeted ribozymes may be used. The term "ribozyme" refers to an RNA-based enzyme capable of targeting and cleaving particular base sequences in oncogene DNA and RNA. Ribozymes either can be targeted directly to cells, in the form of RNA oligo-nucleotides incorporating ribozyme sequences, or introduced into the cell as an expression construct encoding the desired ribozymal RNA. Ribozymes may be used and applied in much the same way as described for antisense nucleic acids.

#### *Methods of Gene Transfer*

In order to mediate the effect of transgene expression in a cell, it will be necessary to transfer the therapeutic expression constructs of the present invention into a cell. This section provides a discussion of methods and compositions of viral production and viral gene transfer, as well as non-viral gene transfer methods.

##### *(i) Viral Vector-Mediated Transfer*

The THAP-family gene is incorporated into a viral infectious particle to mediate gene transfer to a cell. Additional expression constructs encoding other therapeutic agents as described herein may also be transferred via viral transduction using infectious viral particles, for example, by transformation with an adenovirus vector of the present invention as described herein below. Alternatively, retroviral or bovine papilloma virus may be employed, both of which permit permanent transformation of a host cell with a gene(s) of interest. Thus, in one example, viral

infection of cells is used in order to deliver therapeutically significant genes to a cell. Typically, the virus simply will be exposed to the appropriate host cell under physiologic conditions, permitting uptake of the virus. Though adenovirus is exemplified, the present methods may be advantageously employed with other viral or non-viral vectors, as discussed below.

5           **Adenovirus**

Adenovirus is particularly suitable for use as a gene transfer vector because of its mid-sized DNA genome, ease of manipulation, high titer, wide target-cell range, and high infectivity. The roughly 36 kB viral genome is bounded by 100-200 base pair (bp) inverted terminal repeats (ITR), in which are contained cis acting elements necessary for viral DNA replication and packaging. The early (E) and late (L) regions of the genome that contain different transcription units are divided by the onset of viral DNA replication.

The E1 region (E1A and E1B) encodes proteins responsible for the regulation of transcription of the viral genome and a few cellular genes. The expression of the E2 region (E2A and E2B) results in the synthesis of the proteins for viral DNA replication.

15           These proteins are involved in DNA replication, late gene expression, and host cell shut off (Renan, 1990). The products of the late genes (L1, L2, U, L4 and L5), including the majority of the viral capsid proteins, are expressed only after significant processing of a single primary transcript issued by the major late promoter (MLP). The MLP (located at 16.8 map units) is particularly efficient during the late phase of infection, and all the mRNAs issued from this promoter possess a 5' tripartite leader (TL) sequence which makes them preferred mRNAs for translation.

20           In order for adenovirus to be optimized for gene therapy, it is necessary to maximize the carrying capacity so that large segments of DNA can be included. It also is very desirable to reduce the toxicity and immunologic reaction associated with certain adenoviral products. The two goals are, to an extent, coterminous in that elimination of adenoviral genes serves both ends. By practice of the present invention, it is possible achieve both these goals while retaining the ability to manipulate the therapeutic constructs with relative ease.

25           The large displacement of DNA is possible because the cis elements required for viral DNA replication all are localized in the inverted terminal repeats (ITR) (100-200 bp) at either end of the linear viral genome. Plasmids containing ITR's can replicate in the presence of a non-defective adenovirus (Hay et al., 1984). Therefore, inclusion of these elements in an adenoviral vector should permit replication.

30           In addition, the packaging signal for viral encapsidation is localized between 194 385 bp (0.5-1.1 map units) at the left end of the viral genome (Hearing et al., 1987). This signal mimics the protein recognition site in bacteriophage  $\lambda$  DNA where a specific sequence close to the left end, but outside the cohesive end sequence, mediates the binding to proteins that are required for insertion of the DNA into the head structure. E1 substitution vectors of Ad have demonstrated that a 450 bp (0-

1.25 map units) fragment at the left end of the viral genome could direct packaging in 293 cells (Levrero et al., 1991).

Previously, it has been shown that certain regions of the adenoviral genome can be incorporated into the genome of mammalian cells and the genes encoded thereby expressed. These cell lines are capable of supporting the replication of an adenoviral vector that is deficient in the adenoviral function encoded by the cell line. There also have been reports of complementation of replication deficient adenoviral vectors by "helping" vectors, e.g., wild-type virus or conditionally defective mutants.

Replication-deficient adenoviral vectors can be complemented, in trans, by helper virus. This observation alone does not permit isolation of the replication-deficient vectors, however, since the presence of helper virus, needed to provide replicative functions, would contaminate any preparation. Thus, an additional element was needed that would add specificity to the replication and/or packaging of the replication-deficient vector. That element, as provided for in the present invention, derives from the packaging function of adenovirus.

It has been shown that a packaging signal for adenovirus exists in the left end of the conventional adenovirus map (Tibbetts, 1977). Later studies showed that a mutant with a deletion in the EIA (194-358 bp) region of the genome grew poorly even in a cell line that complemented the early (EIA) function (Hearing and Shenk, 1983). When a compensating adenoviral DNA (0-353 bp) was recombined into the right end of the mutant, the virus was packaged normally. Further mutational analysis identified a short, repeated, position-dependent element in the left end of the Ad5 genome. One copy of the repeat was found to be sufficient for efficient packaging if present at either end of the genome, but not when moved towards the interior of the Ad5 DNA molecule (Hearing et al., 1987).

By using mutated versions of the packaging signal, it is possible to create helper viruses that are packaged with varying efficiencies. Typically, the mutations are point mutations or deletions. When helper viruses with low efficiency packaging are grown in helper cells, the virus is packaged, albeit at reduced rates compared to wild-type virus, thereby permitting propagation of the helper. When these helper viruses are grown in cells along with virus that contains wild-type packaging signals, however, the wild-type packaging signals are recognized preferentially over the mutated versions. Given a limiting amount of packaging factor, the virus containing the wild-type signals are packaged selectively when compared to the helpers. If the preference is great enough, stocks approaching homogeneity should be achieved.

### **Retrovirus**

The retroviruses are a group of single-stranded RNA viruses characterized by an ability to convert their RNA to double-stranded DNA in infected cells by a process of reverse-transcription (Coffin, 1990). The resulting DNA then stably integrates into cellular chromosomes as a provirus and directs synthesis of viral proteins.

The integration results in the retention of the viral gene sequences in the recipient cell and its descendants. The retroviral genome contains three genes - gag, pol and env - that code for capsid proteins, polymerase enzyme, and envelope components, respectively. A sequence found upstream from the gag gene, termed T, functions as a signal for packaging of the genome into virions. Two long terminal repeat (LTR) sequences are present at the 5' and 3' ends of the viral genome. These contain strong promoter and enhancer sequences and also are required for integration in the host cell genome (Coffin, 1990).

In order to construct a retroviral vector, a nucleic acid encoding a promoter is inserted into the viral genome in the place of certain viral sequences to produce a virus that is replication-defective. In order to produce virions, a packaging cell line containing the gag, pol and env genes but without the LTR and T components is constructed (Mann et al., 1983). When a recombinant plasmid containing a human cDNA, together with the retroviral LTR and T sequences is introduced into this cell line (by calcium phosphate precipitation for example), the T sequence allows the RNA transcript of the recombinant plasmid to be packaged into viral particles, which are then secreted into the culture media (Nicolas and Rubenstein, 1988; Temin, 1986; Mann et al., 1983). The media containing the recombinant retroviruses is collected, optionally concentrated, and used for gene transfer. Retroviral vectors are able to infect a broad variety of cell types. However, integration and stable expression of many types of retroviruses require the division of host cells (Paskind et al., 1975).

An approach designed to allow specific targeting of retrovirus vectors recently was developed based on the chemical modification of a retrovirus by the chemical addition of galactose residues to the viral envelope. This modification could permit the specific infection of cells such as hepatocytes via asialoglycoprotein receptors, should this be desired.

A different approach to targeting of recombinant retroviruses was designed in which biotinylated antibodies against a retroviral envelope protein and against a specific cell receptor were used. The antibodies were coupled via the biotin components by using streptavidin (Roux et al., 1989). Using antibodies against major histocompatibility complex class I and class II antigens, the infection of a variety of human cells that bore those surface antigens was demonstrated with an ecotropic virus in vitro (Roux et al., 1989).

#### **Adeno-associated Virus**

AAV utilizes a linear, single-stranded DNA of about 4700 base pairs. Inverted terminal repeats flank the genome. Two genes are present within the genome, giving rise to a number of distinct gene products. The first, the cap gene, produces three different virion proteins (VP), designated VP-1, VP 2 and VP-3.

The second, the rep gene, encodes four non-structural proteins (NS). One or more of these rep gene products is responsible for transactivating AAV transcription.

The three promoters in AAV are designated by their location, in map units, in the genome. These are, from left to right, p5, p19 and p40. Transcription gives rise to six transcripts, two initiated at each of three promoters, with one of each pair being spliced.

5 The splice site, derived from map units 42-46, is the same for each transcript. The four non-structural proteins apparently are derived from the longer of the transcripts, and three virion proteins all arise from the smallest transcript.

AAV is not associated with any pathologic state in humans. Interestingly, for efficient replication, AAV requires "helping" functions from viruses such as herpes simplex virus I and II, cytomegalovirus, pseudorabies virus and, of course, adenovirus.

10 The best characterized of the helpers is adenovirus, and many "early" functions for this virus have been shown to assist with AAV replication. Low level expression of AAV rep proteins is believed to hold AAV structural expression in check, and helper virus infection is thought to remove this block.

The terminal repeats of the AAV vector can be obtained by restriction endonuclease digestion of AAV or a plasmid such as p201, which contains a modified AAV genome (Samulski et al, 1987), or by other methods known to the skilled artisan, including but not limited to chemical or enzymatic synthesis of the terminal repeats based upon the published sequence of AAV. The ordinarily skilled artisan can determine, by well-known methods such as deletion analysis, the minimum sequence or part of the AAV ITRs which is required to allow function, i.e., stable and site specific integration.

20 The ordinarily skilled artisan also can determine which minor modifications of the sequence can be tolerated while maintaining the ability of the terminal repeats to direct stable, site-specific integration.

AAV-based vectors have proven to be safe and effective vehicles for gene delivery in vitro, and these vectors are being developed and tested in pre-clinical and clinical stages for a wide range of applications in potential gene therapy, both ex vivo and in vivo (Carter and Flotte, 1996; Chattedee et al., 1995; Ferrari et al., 1996; Fisher et al., 1996; Flotte et al., 1993; Goodman et al., 1994; Kaplitt et al., 1994; 1996, Kessler et al., 1996; Koeberl et al., 1997; Mizukami et al., 1996; Xiao et al., 1996).

30 AAV-mediated efficient gene transfer and expression in the lung has led to clinical trials for the treatment of cystic fibrosis (Carter and Flotte, 1996; Flotte et al., 1993). Similarly, the prospects for treatment of muscular dystrophy by AAV-mediated gene delivery of the dystrophin gene to skeletal muscle, of Parkinson's disease by tyrosine hydroxylase gene delivery to the brain, of hemophilia B by Factor IX gene delivery to the liver, and potentially of myocardial infarction by vascular endothelial growth factor gene to the heart, appear promising since AAV-mediated transgene expression in these organs has recently been shown to be highly efficient (Fisher et al.,

1996; Flotte et al., 1993; Kaplitt et al., 1994; 1996; Koeberl et al., 1997; McCown et al., 1996; Ping et al., 1996; and Xiao et al., 1996).

#### Other Viral Vectors

Other viral vectors may be employed as expression constructs in the present invention.

5 Vectors derived from viruses such as vaccinia virus (Ridgeway, 1988; Baichwal and Sugden, 1986; Coupar et al., 1988) and hepatitis B viruses have also been developed and are useful in the present invention. They offer several attractive features for various mammalian cells (Friedmann, 1989; Ridgeway, 1988; Baichwal and Sugden, 1986; Coupar et al., 1988; and Horwich et al., 1990).

10 With the recent recognition of defective hepatitis B viruses, new insight was gained into the structure-function relationship of different viral sequences. In vitro studies showed that the virus could retain the ability for helper dependent packaging and reverse transcription despite the deletion of up to 80% of its genome (Horwich et al., 1990). This suggested that large portions of the genome could be replaced with foreign genetic material. Chang et al., recently introduced the chloramphenicol acetyltransferase (CAT) gene into duck hepatitis B virus genome in the place of  
15 the polymerase, surface, and pre-surface coding sequences. It was cotransfected with wild-type virus into an avian hepatoma cell line. Culture media containing high titers of the recombinant virus were used to infect primary duckling hepatocytes. Stable CAT gene expression was detected for at least 24 days after transfection (Chang et al., 1991).

In still further embodiments of the present invention, the nucleic acids to be delivered are  
20 housed within an infective virus that has been engineered to express a specific binding ligand. The virus particle will thus bind specifically to the cognate receptors of the target cell and deliver the contents to the cell. A novel approach designed to allow specific targeting of retrovirus vectors was recently developed based on the chemical modification of a retrovirus by the chemical addition of lactose residues to the viral envelope. This modification can permit the specific infection of  
25 hepatocytes via sialoglycoprotein receptors.

Another approach to targeting of recombinant retroviruses was designed in which biotinylated antibodies against a retroviral envelope protein and against a specific cell receptor were used. The antibodies were coupled via the biotin components by using streptavidin (Roux et al., 1989). Using antibodies against major histocompatibility complex class I and class II antigens, they  
30 demonstrated the infection of a variety of human cells that bore those surface antigens with an ecotropic virus in vitro (Roux et al., 1989).

#### *(ii) Non-viral Transfer*

DNA constructs of the present invention are generally delivered to a cell. In certain situations, the nucleic acid to be transferred is non-infectious, and can be transferred using non-viral  
35 methods.

Several non-viral methods for the transfer of expression constructs into cultured mammalian cells are contemplated by the present invention. These include calcium phosphate

precipitation (Graham and Van Der Eb, 1973; Chen and Okayama, 1987; Rippe et al., 1990) DEAE-dextran (Gopal, 1985), electroporation (Tur Kaspas et al., 1986; Potter et al., 1984), direct microinjection (Harland and Weintraub, 1985), DNA loaded liposomes (Nicolau and Sene, 1982; Fraley et al., 1979), cell sonication (Fechheimer et al., 1987), gene bombardment using high velocity microprojectiles (Yang et al., 1990), and receptor-mediated transfection (Wu and Wu, 1987; Wu and Wu, 1988).

Once the construct has been delivered into the cell the nucleic acid encoding the therapeutic gene may be positioned and expressed at different sites. In certain embodiments, the nucleic acid encoding the therapeutic gene may be stably integrated into the genome of the cell. This integration may be in the cognate location and orientation via homologous recombination (gene replacement) or it may be integrated in a random, non-specific location (gene augmentation). In yet further embodiments, the nucleic acid may be stably maintained in the cell as a separate, episomal segment of DNA. Such nucleic acid segments or "episomes" encode sequences sufficient to permit maintenance and replication independent of or in synchronization with the host cell cycle.

How the expression construct is delivered to a cell and where in the cell the nucleic acid remains is dependent on the type of expression construct employed.

In a particular embodiment of the invention, the expression construct may be entrapped in a liposome. Liposomes are vesicular structures characterized by a phospholipid bilayer membrane and an inner aqueous medium. Multilamellar liposomes have multiple lipid layers separated by aqueous medium. They form spontaneously when phospholipids are suspended in an excess of aqueous solution. The lipid components undergo self-rearrangement before the formation of closed structures and entrap water and dissolved solutes between the lipid bilayers (Ghosh and Bachhawat, 1991). The addition of DNA to cationic liposomes causes a topological transition from liposomes to optically birefringent liquid-crystalline condensed globules (Radler et al., 1997). These DNA-lipid complexes are potential non-viral vectors for use in gene therapy.

Liposome-mediated nucleic acid delivery and expression of foreign DNA in vitro has been very successful. Using the P-lactamase gene, Wong et al. (1980) demonstrated the feasibility of liposome-mediated delivery and expression of foreign DNA in cultured chick embryo, HeLa, and hepatoma cells. Nicolau et al. (1987) accomplished successful liposome-mediated gene transfer in rats after intravenous injection. Also included are various commercial approaches involving "lipofection" technology.

In certain embodiments of the invention, the liposome may be complexed with a hemagglutinating virus (HVJ). This has been shown to facilitate fusion with the cell membrane and promote cell entry of liposome-encapsulated DNA (Kaneda et al., 1989).

In other embodiments, the liposome may be complexed or employed in conjunction with nuclear nonhistone chromosomal proteins (HMG-1) (Kato et al., 1991). In yet further embodiments, the liposome may be complexed or employed in conjunction with both HVJ and HMG-1. In that

such expression constructs have been successfully employed in transfer and expression of nucleic acid in vitro and in vivo, then they are applicable for the present invention.

Other vector delivery systems which can be employed to deliver a nucleic acid encoding a therapeutic gene into cells are receptor-mediated delivery vehicles. These take advantage of the selective uptake of macromolecules by receptor mediated endocytosis in almost all eukaryotic cells. Because of the cell type specific distribution of various receptors, the delivery can be highly specific (Wu and Wu, 1993).

Receptor-mediated gene targeting vehicles generally consist of two components: a cell receptor-specific ligand and a DNA-binding agent. Several ligands have been used for receptor-mediated gene transfer. The most extensively characterized ligands are asialoorosomucoid (ASOR) (Wu and Wu, 1987) and transferring (Wagner et al., 1990).

Recently, a synthetic neoglycoprotein, which recognizes the same receptor as ASOR, has been used as a gene delivery vehicle (Ferkol et al., 1993; Perales et al., 1994) and epidermal growth factor (EGF) has also been used to deliver genes to squamous carcinoma cells (Myers, EPO 0273085).

In other embodiments, the delivery vehicle may comprise a ligand and a liposome. For example, Nicolau et al, (1987) employed lactosyl-ceramide, a galactose terminal asialganglioside, incorporated into liposomes and observed an increase in the uptake of the insulin gene by hepatocytes. Thus, it is feasible that a nucleic acid encoding a therapeutic gene also may be specifically delivered into a cell type such as prostate, epithelial or tumor cells, by any number of receptor-ligand systems with or without liposomes. For example, the human prostate-specific antigen (Watt et al, 1986) may be used as the receptor for mediated delivery of a nucleic acid in prostate tissue.

In another embodiment of the invention, the expression construct may simply consist of naked recombinant DNA or plasmids. Transfer of the construct may be performed by any of the methods mentioned above which physically or chemically permeabilize the cell membrane. This is applicable particularly for transfer in vitro, however, it may be applied for in vivo use as well. Dubensky et al, (1984) successfully injected polyornavirus DNA in the form of CaP04 precipitates into liver and spleen of adult and newborn mice demonstrating active viral replication and acute infection.

Benvenisty and Neshif (1986) also demonstrated that direct intraperitoneal injection of CaP04 precipitated plasmids results in expression of the transfected genes. It is envisioned that DNA encoding a CAM may also be transferred in a similar manner in vivo and express CAM.

Another embodiment of the invention for transferring a naked DNA expression construct into cells may involve particle bombardment. This method depends on the ability to accelerate DNA coated microprojectiles to a high velocity allowing them to pierce cell membranes and enter cells without killing them (Klein et al, 1987). Several devices for accelerating small particles have



been developed. One such device relies on a high voltage discharge to generate an electrical current, which in turn provides the motive force (Yang et al, 1990). The microprojectiles used have consisted of biologically inert substances such as tungsten or gold beads.

#### Antibodies

5 Polyclonal anti-THAP-family or anti-THAP domain antibodies can be prepared as described above by immunizing a suitable subject with a THAP-family or THAP domain immunogen. The anti-THAP-family or anti-THAP domain antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized THAP-family or THAP domain protein. If desired, the antibody  
10 molecules directed against THAP-family can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, e.g., when the anti-THAP-family antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as those described in the following references:  
15 the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497) (see also, Brown et al. (1981) *J. Immunol.* 127:539-46; Brown et al. (1980) *J. Biol. Chem.* 255:4980-83 ; Yeh et al. (1976) *PNAS* 76:2927-31; and Yeh et al. (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor et al. (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan  
20 R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally R. H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, N.Y. (1980); E. A. Lerner (1981) *Yale J. Biol. Med.*, 54:387-402; M. L. Gefter et al. (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes)  
25 from a mammal immunized with a THAP-family immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds THAP-family.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-THAP-family or anti-THAP domain  
30 monoclonal antibody (see, e.g., G. Galfre et al. (1977) *Nature* 266:55052; Gefter et al. *Somatic Cell Genet.*, cited supra; Lerner, *Yale J Biol. Med*, cited supra; Kenneth, *Monoclonal Antibodies*, cited supra). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (e.g., a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine  
35 hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine,

aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, e.g., the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG").

5 Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind a THAP-family or THAP domain protein, e.g., using a standard ELISA assay.

10 Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-THAP-family or anti-THAP domain antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with THAP-family or THAP domain protein to thereby isolate immunoglobulin library members that bind THAP-family or THAP domain proteins. Kits for generating and screening phage display  
15 libraries are commercially available (e.g., the Pharmacia Recombinant Phage Antibody System, Catalog No. 27-9400-01; and the Stratagene SurfZAP.TM. Phage Display Kit, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner et al. U.S. Pat. No. 5,223,409; Kang et al. PCT International Publication No. WO 92/18619; Dower et al. PCT  
20 International Publication No. WO 91/17271; Winter et al. PCT International Publication WO 92/20791; Markland et al. PCT International Publication No. WO 92/15679; Breitling et al. PCT International Publication WO 93/01288; McCafferty et al. PCT International Publication No. WO 92/01047; Garrard et al. PCT International Publication No. WO 90/02809; Fuchs et al. (1991) Bio/Technology 9:1370-1372; Hay  
25 et al. (1992) Hum. Antibod. Hybridomas 3:81-85; Huse et al. (1989) Science 246:1275-1281; Griffiths et al. (1993) EMBO J 12:725-734; Hawkins et al. (1992) J. Mol. Biol. 226:889-896; Clarkson et al. (1991) Nature 352:624-628; Gram et al. (1992) PNAS 89:3576-3580; Garrad et al. (1991) Bio/Technology 9:1373-1377; Hoogenboom et al. (1991) Nuc. Acid Res. 19:4133-4137; Barbas et al. (1991) PNAS 88:7978-7982; and McCafferty et al. Nature (1990) 348:552-554.

30 Additionally, recombinant anti-THAP-family or anti-THAP domain antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson et al.  
35 International Application No. PCT/US86/02269; Akira, et al. European Patent Application 184,187; Taniguchi, M., European Patent Application 171496; Morrison et al. European Patent Application 173,494; Neuberger et al. PCT International Publication No. WO 86/01533; Cabilly et al. U.S. Pat.

No. 4,816,567; Cabilly et al. European Patent Application 125,023; Better et al. (1988) Science 240:1041-1043; Liu et al. (1987) PNAS 84:3439-3443; Liu et al. (1987) J. Immunol. 139:3521-3526; Sun et al. (1987) PNAS 84:214-218; Nishimura et al. (1987) Canc. Res. 47:999-1005; Wood et al. (1985) Nature 314:446-449; and Shaw et al. (1988) J. Natl. Cancer Inst. 80:1553-1559);  
5 Morrison, S. L. (1985) Science 229:1202-1207; Oi et al. (1986) BioTechniques 4:214; Winter U.S. Pat. No. 5,225,539; Jones et al. (1986) Nature 321:552-525; Verhoeyan et al. (1988) Science 239:1534; and Beidler et al. (1988) J. Immunol. 141:4053-4060.

An anti-THAP-family of anti-THAP domain antibody (e.g., monoclonal antibody) can be used to isolate THAP-family or THAP domain protein by standard techniques, such as affinity  
10 chromatography or immunoprecipitation. For example, an anti-THAP-family antibody can facilitate the purification of natural THAP-family from cells and of recombinantly produced THAP-family expressed in host cells. Moreover, an anti-THAP-family antibody can be used to detect THAP-family protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the THAP-family protein. Anti-THAP-family antibodies can be used  
15 diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes  
20 include horseradish peroxidase, alkaline phosphatase, -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include  
25 luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

#### DRUG SCREENING ASSAYS

Some embodiments of the present invention provide a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g.,  
30 preferably small molecules, but also peptides, peptidomimetics or other drugs) which bind to THAP-family or THAP domain proteins, have an inhibitory or activating effect on, for example, THAP-family expression or preferably THAP-family activity, or have an inhibitory or activating effect on, for example, the activity of an THAP-family target molecule. In some embodiments small molecules can be generated using combinatorial chemistry or can be obtained from a natural  
35 products library. Assays may be cell based, non-cell-based or in vivo assays. Drug screening assays may be binding assays or more preferentially functional assays, as further described.

In general, any suitable activity of a THAP-family protein can be detected in a drug screening assay, including: (1) mediating apoptosis or cell proliferation when expressed or introduced into a cell, most preferably inducing or enhancing apoptosis, and/or most preferably reducing cell proliferation; (2) mediating apoptosis or cell proliferation of an endothelial cell; (3) mediating apoptosis or cell proliferation of a hyperproliferative cell; (4) mediating apoptosis or cell proliferation of a CNS cell, preferably a neuronal or glial cell; (5) an activity indicative of a biological function in an animal selected from the group consisting of mediating, for example enhancing or inhibiting, angiogenesis; mediating, preferably inhibiting, inflammation; inhibiting the metastatic potential of cancerous tissue; reducing tumor burden; increasing sensitivity of cancerous cells to chemotherapy or radiotherapy; killing a cancer cell, inhibiting the growth of a cancer cell, inducing tumor regression; and mediating, preferably inhibiting, one or more of the following conditions, T-cell auto-immune infiltrative skin diseases, chronic autoinflammatory skin diseases, such as lichen panus and psoriasis, autoimmune encephalomyelitis, multiple sclerosis, rheumatoid arthritis, autoimmune diabetes, inflammatory bowel diseases, such as Crohn's disease and ulcerative colitis, Hashimoto's thyroiditis, Sjogren's syndrome, gastric lymphomas and chronic inflammatory liver disease or (6) interaction with a THAP family target molecule or THAP domain target molecule, preferably interaction with a protein or a nucleic acid.

The invention also provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g., preferably small molecules, but also peptides, peptidomimetics or other drugs) which bind to THAP1, PAR4 or PML-NB proteins, and have an inhibitory or activating effect on PAR4 or THAP1 recruitment, binding to or association with PML-NBs or interaction of a chemokine with a THAP-family polypeptide or a cellular response to a chemokine which is mediated by a THAP-family polypeptide.

In one embodiment, the invention provides assays for screening candidate or test compounds which are target molecules of a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is used with peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K. S. (1997) Anticancer Drug Des. 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al. (1993) Proc. Natl. Acad. Sci. U.S.A. 90:6909; Erb et al. (1994) Proc. Natl. Acad. Sci. USA 91:11422; Zuckermann et al. (1994). J. Med. Chem. 37:2678; Cho et al. (1993) Science 261:1303; Carrell et al. (1994) Angew. Chem. Int. Ed. Engl. 33:2059; Carrell et al. (1994) Angew. Chem. Int. Ed. Engl. 33:2061; and in Gallop et al. (1994) J. Med. Chem. 37:1233.

Libraries of compounds may be presented in solution (e.g., Houghten (1992) Biotechniques 13:412-421), or on beads (Lam (1991) Nature 354:82-84), chips (Fodor (1993) Nature 364:555-556), bacteria (Ladner U.S. Pat. No. 5,223,409), spores (Ladner U.S. Pat. No. '409), plasmids (Cull et al. (1992) Proc Natl Acad Sci USA 89:1865-1869) or on phage (Scott and Smith (1990) Science 249:386-390); (Devin (1990) Science 249:404-406); (Cwirla et al. (1990) Proc. Natl. Acad. Sci. 87:6378-6382); (Felici (1991) J. Mol. Biol. 222:301-310); (Ladner supra.).

Determining the ability of the test compound to inhibit or increase THAP-family polypeptide activity can also be accomplished, for example, by coupling the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof with a radioisotope or enzymatic label such that binding of the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof to its cognate target molecule can be determined by detecting the labeled THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof in a complex. For example, compounds (e.g., THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof) can be labeled with  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ , or  $^3\text{H}$ , either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. The labeled molecule is placed in contact with its cognate molecule and the extent of complex formation is measured. For example, the extent of complex formation may be measured by immunoprecipitating the complex or by performing gel electrophoresis.

It is also within the scope of this invention to determine the ability of a compound (e.g., THAP family or THAP domain polypeptide, or biologically active fragment or homologue thereof) to interact with its cognate target molecule without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a compound with its cognate target molecule without the labeling of either the compound or the target molecule. McConnell, H. M. et al. (1992) Science 257:1906-1912. A microphysiometer such as a cytosensor is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between compound and cognate target molecule.

In a preferred embodiment, the assay comprises contacting a cell which expresses a THAP family or THAP domain polypeptide, or biologically active fragment or homologue thereof, with a

THAP-family or THAP domain protein target molecule to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to inhibit or increase the activity of the THAP family or THAP domain polypeptide, or biologically active fragment or homologue thereof, wherein determining the ability of the test compound to inhibit or increase the activity of the THAP family or THAP domain polypeptide, or biologically active fragment or homologue thereof, comprises determining the ability of the test compound to inhibit or increase a biological activity of the THAP-family polypeptide expressing cell.

In another embodiment, the assay comprises contacting a cell which expresses a THAP family or THAP domain polypeptide, or biologically active fragment or homologue thereof, with a test compound, and determining the ability of the test compound to inhibit or increase the activity of the THAP family or THAP domain polypeptide, or biologically active fragment or homologue thereof, wherein determining the ability of the test compound to inhibit or increase the activity of the THAP family or THAP domain polypeptide, or biologically active fragment or homologue thereof, comprises determining the ability of the test compound to inhibit or increase a biological activity of the THAP-family polypeptide expressing cell.

In another preferred embodiment, the assay comprises contacting a cell which is responsive to a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof, with a THAP-family protein or biologically-active portion thereof, to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to modulate the activity of the THAP-family protein or biologically active portion thereof, wherein determining the ability of the test compound to modulate the activity of the THAP-family protein or biologically active portion thereof comprises determining the ability of the test compound to modulate a biological activity of the THAP-family polypeptide-responsive cell (e.g., determining the ability of the test compound to modulate a THAP-family polypeptide activity).

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a THAP-family target molecule (i.e. a molecule with which THAP-family polypeptide interacts) with a test compound and determining the ability of the test compound to modulate (e.g. stimulate or inhibit) the activity of the THAP-family target molecule. Determining the ability of the test compound to modulate the activity of a THAP-family target molecule can be accomplished, for example, by determining the ability of the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof to bind to or interact with the THAP-family target molecule.

Determining the ability of the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof to bind to or interact with a THAP-family target molecule can be accomplished by one of the methods described above for determining direct binding. In a preferred embodiment, determining the ability of the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof to bind to or interact with a THAP-family

target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by contacting the target molecule with the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof and measuring induction of a cellular second messenger of the target (i.e. intracellular  $\text{Ca}^{2+}$ , diacylglycerol,  $\text{IP}_3$ , etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a target-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, e.g., luciferase), or detecting a target-regulated cellular response, for example, signal transduction or protein:protein interactions.

10 In yet another embodiment, an assay of the present invention is a cell-free assay in which a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof is contacted with a test compound and the ability of the test compound to bind to the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof is determined. Binding of the test compound to the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof with a known compound which binds THAP-family polypeptide (e.g., a THAP-family target molecule) to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof, wherein determining the ability of the test compound to interact with a THAP-family protein comprises determining the ability of the test compound to preferentially bind to THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof as compared to the known compound.

25 In another embodiment, the assay is a cell-free assay in which a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof is contacted with a test compound and the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof is determined. Determining the ability of the test compound to modulate the activity of a THAP-family protein can be accomplished, for example, by determining the ability of the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof to bind to a THAP-family target molecule by one of the methods described above for determining direct binding. Determining the ability of the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof to bind to a THAP-family target molecule can also be accomplished using a technology such as real-time Biomolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) Anal. Chem. 63:2338-2345 and Szabo et al. (1995) Curr. Opin. Struct. Biol. 5:699-705. As used herein, "BIA" is a technology for studying biospecific

interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the optical phenomenon of surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

5 In an alternative embodiment, determining the ability of the test compound to modulate the activity of a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof can be accomplished by determining the ability of the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof to further modulate the activity of a downstream effector (e.g., a growth factor mediated signal transduction pathway component) of a THAP-family target molecule. For example, the activity of the effector molecule  
10 on an appropriate target can be determined or the binding of the effector to an appropriate target can be determined as previously described.

In yet another embodiment, the cell-free assay involves contacting a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof with a known compound which binds the THAP-family protein to form an assay mixture, contacting the assay  
15 mixture with a test compound, and determining the ability of the test compound to interact with the THAP-family protein, wherein determining the ability of the test compound to interact with the THAP-family protein comprises determining the ability of the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof to preferentially bind to or modulate the activity of a THAP-family target molecule.

20 The cell-free assays of the present invention are amenable to use of both soluble and/or membrane-bound forms of isolated proteins (e.g. THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof or molecules to which THAP-family targets bind). In the case of cell-free assays in which a membrane-bound form of an isolated protein is used it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the isolated  
25 protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton.RTM. X-100, Triton.RTM. X-114, Thesit.RTM., Isotridecypoly(ethylene glycol ether)<sub>n</sub>, 3-[(3-cholamidopropyl)dimethylamminio]- 1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate  
30 (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof or a target molecule thereof to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate  
35 automation of the assay. Binding of a test compound to a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof, or interaction of a THAP-family protein with a target molecule in the presence and absence of a candidate compound, can be



accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/THAP-family fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, Mo.) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or THAP-family protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of THAP-family polypeptide binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either a THAP-family protein or a THAP-family target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated THAP-family protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with a THAP-family protein or target molecule but which do not interfere with binding of the THAP-family protein to its target molecule can be derivatized to the wells of the plate, and unbound target or THAP-family protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the THAP-family protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the THAP-family protein or target molecule.

In another embodiment, modulators of THAP-family or THAP domain polypeptides expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of THAP-family or THAP domain polypeptides mRNA or protein in the cell is determined. The level of expression of THAP-family polypeptide mRNA or protein in the presence of the candidate compound is compared to the level of expression of THAP-family polypeptide or THAP domain mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of THAP-family polypeptide expression based on this comparison. For example, when expression of THAP-family polypeptide or THAP domain mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of THAP-family polypeptide or THAP domain mRNA or protein expression. Alternatively, when expression of

THAP-family polypeptide or THAP domain mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of THAP-family polypeptide or THAP domain mRNA or protein expression. The level of THAP-family polypeptide or THAP domain mRNA or protein expression  
5 in the cells can be determined by methods described herein for detecting THAP-family polypeptide or THAP domain mRNA or protein.

In yet another aspect of the invention, the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay using the methods described above for use in THAP-family  
10 polypeptide/PAR4 interactions assays, to identify other proteins which bind to or interact with THAP-family polypeptide ("THAP-family-binding proteins" or "THAP-family-bp") and are involved in THAP-family polypeptide activity. Such THAP-family- or THAP domain-binding proteins are also likely to be involved in the propagation of signals by the THAP-family or THAP domain proteins or THAP-family or THAP domain proteins targets as, for example, downstream  
15 elements of a THAP-family polypeptide- or THAP domain-mediated signaling pathway. Alternatively, such THAP-family-binding proteins are likely to be THAP-family polypeptides inhibitors.

#### THAP/DNA BINDING ASSAYS

In another embodiment of the invention a method is provided for identifying compounds  
20 which interfere with THAP-family DNA binding activity, comprising the steps of: contacting a THAP-family protein or a portion thereof immobilized on a solid support with both a test compound and DNA fragments, or contacting a DNA fragment immobilized on a solid support with both a test compound and a THAP-family protein. The binding between DNA and the THAP-protein or a portion thereof is detected, wherein a decrease in DNA binding when compared to  
25 DNA binding in the absence of the test compound indicates that the test compound is an inhibitor of THAP-family DNA binding activity, and an increase in DNA binding when compared to DNA binding in the absence of the test compound indicates that the test compound is an inducer of or restores THAP-family DNA binding activity. As discussed further, DNA fragments may be selected to be specific THAP-family protein target DNA obtained for example as described in  
30 Example 28, or may be non-specific THAP-family target DNA. Methods for detecting protein-DNA interactions are well known in the art, including most commonly used electrophoretic mobility shift assays (EMSAs) or by filter binding (Zabel et al, (1991) J. Biol. Chem., 266:252; and Okamoto and Beach, (1994) Embo J. 13: 4816). Other assays are available which are amenable for high throughput detection and quantification of specific and nonspecific DNA binding (Amersham,  
35 N.J.; and Gal S. et al, 6<sup>th</sup> Ann. Conf. Soc. Biomol. Screening, 6-9 Sept 2000, Vancouver, B.C.).

In a first aspect, a screening assay involves identifying compounds which interfere with THAP-family DNA binding activity without prior knowledge about specific THAP-family binding

sequences. For example, a THAP-family protein is contacted with both a test compound and a library of oligonucleotides or a sample of DNA fragments not selected based on specific DNA sequences. Preferably the THAP-family protein is immobilized on a solid support (such as an array or a column). Unbound DNA is separated from DNA which is bound to the THAP-family protein, and the DNA which is bound to THAP-family protein is detected and can be quantitated by any means known in the art. For example, the DNA fragment is labelled with a detectable moiety, such as a radioactive moiety, a colorimetric moiety or a fluorescent moiety. Techniques for so labelling DNA are well known in the art.

The DNA which is bound to the THAP-family protein or a portion thereof is separated from unbound DNA by immunoprecipitation with antibodies which are specific for the THAP-family protein or a portion thereof. Use of two different monoclonal anti-THAP-family antibodies may result in more complete immunoprecipitation than either one alone. The amount of DNA which is in the immunoprecipitate can be quantitated by any means known in the art. THAP-family proteins or portions thereof which bind to the DNA can also be detected by gel shift assays (Tan, Cell, 62:367, 1990), nuclease protection assays, or methylase interference assays.

It is still another object of the invention to provide methods for identifying compounds which restore the ability of mutant THAP-family proteins or portions thereof to bind to DNA sequences. In one embodiment a method of screening agents for use in therapy is provided comprising: measuring the amount of binding of a THAP-family protein or a portion thereof which is encoded by a mutant gene found in cells of a patient to DNA molecules, preferably random oligonucleotides or DNA fragments from a nucleic acid library; measuring the amount of binding of said THAP-family protein or a portion thereof to said nucleic acid molecules in the presence of a test substance; and comparing the amount of binding of the THAP-family protein or a portion thereof in the presence of said test substance to the amount of binding of the THAP-family protein in the absence of said test substance, a test substance which increases the amount of binding being a candidate for use in therapy.

In another embodiment of the invention, oligonucleotides can be isolated which restore to mutant THAP-family proteins or portions thereof the ability to bind to a consensus binding sequence or conforming sequences. Mutant THAP-family protein or a portion thereof and random oligonucleotides are added to a solid support on which THAP-family-specific DNA fragments are immobilized. Oligonucleotides which bind to the solid support are recovered and analyzed. Those whose binding to the solid support is dependent on the presence of the mutant THAP-family protein are presumptively binding the support by binding to and restoring the conformation of the mutant protein.

If desired, specific binding can be distinguished from non-specific binding by any means known in the art. For example, specific binding interactions are stronger than non-specific binding interactions. Thus the incubation mixture can be subjected to any agent or condition which

destabilizes protein/DNA interactions such that the specific binding reaction is the predominant one detected. Alternatively, as taught more specifically below, a non-specific competitor, such as dI-dC, can be added to the incubation mixture. If the DNA containing the specific binding sites is labelled and the competitor is unlabeled, then the specific binding reactions will be the ones predominantly detected upon measuring labelled DNA.

According to another embodiment of the invention, after incubation of THAP-family protein or a portion thereof with specific DNA fragments all components of the cell lysate which do not bind to the DNA fragments are removed. This can be accomplished, among other ways, by employing DNA fragments which are attached to an insoluble polymeric support such as agarose, cellulose and the like. After binding, all non-binding components can be washed away, leaving THAP-family protein or a portion thereof bound to the DNA/solid support. The THAP-family protein or a portion thereof can be quantitated by any means known in the art. It can be determined using an immunological assay, such as an ELISA, RIA or Western blotting.

In another embodiment of the invention a method is provided for identifying compounds which specifically bind to THAP-family-specific-DNA sequences, comprising the steps of: contacting a THAP-family-specific DNA fragment immobilized on a solid support with both a test compound and wild-type THAP-family protein or a portion thereof to bind the wild-type THAP-family protein or a portion thereof to the DNA fragment; determining the amount of wild-type THAP-family protein which is bound to the DNA fragment, inhibition of binding of wild-type THAP-family protein by the test compound with respect to a control lacking the test compound suggesting binding of the test compound to the THAP-family-specific DNA binding sequences.

It is still another object of the invention to provide methods for identifying compounds which restore the ability of mutant THAP-family proteins or portions thereof to bind to specific DNA binding sequences. In one embodiment a method of screening agents for use in therapy is provided comprising: measuring the amount of binding of a THAP-family protein or a portion thereof which is encoded by a mutant gene found in cells of a patient to a DNA molecule which comprises more than one monomer of a specific THAP-family target nucleotide sequence; measuring the amount of binding of said THAP-family protein to said nucleic acid molecule in the presence of a test substance; and comparing the amount of binding of the THAP-family protein in the presence of said test substance to the amount of binding of the THAP-family protein or a portion thereof in the absence of said test substance, a test substance which increases the amount of binding being a candidate for use in therapy.

In another embodiment of the invention a method is provided for screening agents for use in therapy comprising: contacting a transfected cell with a test substance, said transfected cell containing a THAP-family protein or a portion thereof which is encoded by a mutant gene found in cells of a patient and a reporter gene construct comprising a reporter gene which encodes an assayable product and a sequence which conforms to a THAP-family DNA binding site, wherein

said sequence is upstream from and adjacent to said reporter gene; and determining whether the amount of expression of said reporter gene is altered by the test substance, a test substance which alters the amount of expression of said reporter gene being a candidate for use in therapy.

In still another embodiment a method of screening agents for use in therapy is provided comprising: adding RNA polymerase ribonucleotides and a THAP-family protein or a portion thereof to a transcription construct, said transcription construct comprising a reporter gene which encodes an assayable product and a sequence which conforms to a THAP-family consensus binding site, said sequence being upstream from and adjacent to said reporter gene, said step of adding being effected in the presence and absence of a test substance; determining whether the amount of transcription of said reporter gene is altered by the presence of said test substance, a test substance which alters the amount of transcription of said reporter gene being a candidate for use in therapy.

According to the present invention compounds which have THAP-family activity are those which specifically complex with a THAP-family-specific DNA binding site. Oligonucleotides and oligonucleotide containing nucleotide analogs are also contemplated among those compounds which are able to complex with a THAP-family-specific DNA binding site.

#### **Further assays to modulate THAP-family polypeptide activity in vivo**

It will be appreciated that any suitable assay that allows detection of THAP-family polypeptide or THAP domain activity can be used. Examples of assays for testing protein interaction, nucleic acid binding or modulation of apoptosis in the presence or absence of a test compound are further described herein. Thus, the invention encompasses a method of identifying a candidate THAP-family polypeptide modulator (e.g. activator or inhibitor), said method comprising:

- a) providing a cell comprising a THAP family or THAP domain polypeptide, or a biologically active fragment or homolog thereof;
- b) contacting said cell with a test compound;
- c) determining whether said compound selectively modulates (e.g. activates or inhibits) THAP-family polypeptide activity, preferably pro-apoptotic activity, or THAP family or THAP domain target binding; wherein a determination that said compound selectively modulates (e.g. activates or inhibits) the activity of said polypeptide indicates that said compound is a candidate modulator (e.g. activator or inhibitor respectively) of said polypeptide. Preferably, the THAP family or THAP domain target is a protein or nucleic acid.

Preferably the cell is a cell which has been transfected with an recombinant expression vector encoding a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof.

Several examples of assays for the detection of apoptosis are described herein, in the section titled "Apoptosis assays". Several examples of assays for the detection of THAP family or

THAP domain target interactions are described herein, including assays for detection of protein interactions and nucleic acid binding.

In one example of an assay for apoptosis activity, a high throughput screening assay for molecules that abrogate or stimulate THAP-family polypeptide proapoptotic activity is provided based on serum-withdrawal induced apoptosis in a 3T3 cell line with tetracycline-regulated expression of a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof. Apoptotic cells can be detected by TUNEL labeling in 96- or 384-wells microplates. A drug screening assay can be carried out along the lines as described in Example 23. 3T3 cells, which have previously been used to analyze the pro-apoptotic activity of PAR4 (Diaz-Meco et al, 1996; Berra et al., 1997), can be transfected with expression vectors encoding a THAP-family or THAP domain polypeptide allowing the ectopic expression of THAP-family polypeptide. Then, the apoptotic response to serum withdrawal is assayed in the presence of a test compound, allowing the identification of test compounds that either enhance or inhibit the ability of THAP-family or THAP domain polypeptide to induce apoptosis. Transfected cells are deprived of serum and cells with apoptotic nuclei are counted. Apoptotic nuclei can be counted by DAPI staining and in situ TUNEL assays.

#### **Further THAP-family polypeptide/THAP-target interaction assays**

In exemplary methods THAP/THAP target interaction assays are described in the context of THAP1 and the THAP target Par4. However, it will be appreciated that assays for screening for modulators of other THAP family members or THAP domains and other THAP target molecules may be carried out by substituting these for THAP1 and Par4 in the methods below. For example, in some embodiments, modulators which affect the interaction between a THAP-family polypeptide and SLC are identified. It will be appreciated, however, that the same assays can be used to determine the interaction between any THAP-target polypeptide (for example, a chemokine) and a THAP-family polypeptide, which comprises an interaction domain for the chemokine. THAP-family polypeptides that can be used in these assays include the polypeptides of SEQ ID NOs: 1-114, biologically active fragments thereof, THAP-family polypeptide oligomers, oligomers comprising a THAP-family chemokine-binding domain, THAP-family polypeptide-immunoglobulin fusions, THAP-family chemokine-binding domain-immunoglobulin fusions and polypeptide homologs having at least 30% amino acid identity to any one of the aforementioned polypeptides.

As demonstrated in Examples 4, 5, 6, and 7 and Figures 3, 4 and 5, the inventors have demonstrated using several experimental methods that THAP1 interacts with the pro-apoptotic protein Par4. In particular, it has been shown that THAP1 interacts with Par4 wild type (Par4) and a Par4 death domain (Par4DD) in a yeast two-hybrid system. Yeast cells were cotransformed with BD7-THAP1 and AD7-Par4, AD7, AD7-Par4DD or AD7-Par4) expression vectors. Transformants

were selected on media lacking histidine and adenine. Identical results were obtained by cotransformation of AD7-THAP1 with BD7-Par4, BD7, BD7-Par4DD or BD7-Par4).

The inventors have also demonstrated *in vitro* binding of THAP1 to GST-Par4DD. Par4DD was expressed as a GST fusion protein, purified on glutathione sepharose and employed as an affinity matrix for binding of *in vitro* translated <sup>35</sup>S-methionine labeled THAP1. GST served as negative control.

Furthermore, the inventors have shown that THAP1 interacts with both Par4DD and SLC *in vivo*. Myc-Par4DD and GFP-THAP1 expression vectors were cotransfected in primary human endothelial cells. Myc-Par4DD was stained with monoclonal anti-myc antibody. Green fluorescence, GFP-THAP1; red fluorescence, Par4DD.

The invention thus encompasses assays for the identification of molecules that modulate (stimulate or inhibit) THAP-family polypeptide/PAR4 binding. In preferred embodiments, the invention includes assays for the identification of molecules that modulate (stimulate or inhibit) THAP1 /PAR4 binding or THAP1/SLC binding.

Four examples of high throughput screening assays include:

- 1) a two hybrid-based assay in yeast to find drugs that disrupt interaction of the THAP-family bait with the PAR4 or SLC as prey
- 2) an *in vitro* interaction assay using recombinant THAP-family polypeptide and PAR4 or SLC proteins
- 3) a chip-based binding assay using recombinant THAP-family polypeptide and PAR4 or SLC proteins
- 2) a fluorescence resonance energy transfer (FRET) cell-based assay using THAP-family polypeptide and PAR4 or SLC proteins fused with fluorescent proteins

The invention thus encompasses a method of identifying a candidate THAP-family polypeptide/PAR4 or SLC interaction modulator, said method comprising:

- a) providing a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof and a PAR4 or SLC polypeptide or fragment thereof;
- b) contacting said THAP family or THAP domain polypeptide with a test compound; and
- c) determining whether said compound selectively modulates (e.g. activates or inhibits) THAP-family/PAR4 or SLC interaction activity.

Also envisioned is a method comprising:

- a) providing a cell comprising a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof and a PAR4 or SLC polypeptide or fragment thereof;
- b) contacting said cell with a test compound; and
- c) determining whether said compound selectively modulates (e.g. activates or inhibits) THAP-family/PAR4 or SLC interaction activity.

In general, any suitable assay for the detection of protein-protein interaction may be used.

In one example, a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof can be used as a "bait protein" and a PAR4 or SLC protein can be used as a "prey protein" (or vice-versa) in a two-hybrid assay (see, e.g., U.S. Pat. No. 5,283,317; Zervos et al. (1993) Cell 72:223-232; Madura et al. (1993) J. Biol. Chem. 268:12046-12054; Bartel et al. (1993) Biotechniques 14:920-924; Iwabuchi et al. (1993) Oncogene 8:1693-1696; and Brent WO94/10300). The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof -is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, the gene that codes for a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, in vivo, forming a THAP-family polypeptide/PAR4 complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the THAP-family protein. This assay can thus be carried out in the presence or absence of a test compound, whereby modulation of THAP-family polypeptide/PAR4 or SLC interaction can be detected by lower or lack of transcription of the reported gene.

In other examples, in vitro THAP-family polypeptide/PAR4 or SLC interaction assays can be carried out, several examples of which are further described herein. For example, a recombinant THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof is contacted with a recombinant PAR4 or SLC protein or biologically active portion thereof, and the ability of the PAR4 or SLC protein to bind to the THAP-family protein is determined. Binding of the PAR4 or SLC protein compound to the THAP-family protein can be determined either directly or indirectly as described herein. In a preferred embodiment, the assay includes contacting the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof with a PAR4 or SLC protein which binds a THAP-family protein (e.g., a THAP-family target molecule) to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a THAP-family protein, wherein determining the ability of the test compound to interact with a THAP-family protein comprises determining the ability of the test compound to preferentially bind to THAP-family or biologically active portion thereof as compared to the PAR4 or SLC protein. For example, the step of



determining the ability of the test compound to interact with a THAP-family protein may comprise determining the ability of the compound to displace Par4 or SLC from a THAP-family protein/Par4 or SLC complex thereby forming a THAP-family protein/compound complex. Alternatively, it will be appreciated that it is also possible to determine the ability of the test compound to interact with a PAR4 or SLC protein, wherein determining the ability of the test compound to interact with a PAR4 or SLC protein comprises determining the ability of the test compound to preferentially bind to PAR4 or SLC or biologically active portion thereof as compared to the THAP-family protein. For example, the step of determining the ability of the test compound to interact with a THAP-family protein may comprise determining the ability of the compound to displace Par4 or SLC from a THAP-family protein/Par4 or SLC complex thereby forming a THAP-family protein/compound complex.

*Assays to modulate THAP-family polypeptide and/or Par4 trafficking in the PML nuclear bodies (PML NBs)*

As demonstrated in Examples 8 and 9, the inventors have demonstrated using several experimental methods that THAP1 and Par4 localize in PML NBs.

The inventors demonstrated that THAP1 is a novel protein associated with PML-nuclear bodies. Double immunofluorescence staining showed colocalization of THAP1 with PML-NBs proteins, PML and Daxx. Primary human endothelial cells were transfected with GFP-THAP1 expression vector; endogenous PML and Daxx were stained with monoclonal anti-PML and polyclonal anti-Daxx antibodies, respectively.

The inventors also demonstrated that Par4 is a novel component of PML-NBs that colocalizes with THAP1 *in vivo* by several experiments. In one experiments, double immunofluorescence staining revealed colocalization of Par4 and PML at PML-NBs in primary human endothelial cells or fibroblasts. Endogenous PAR4 and PML were stained with polyclonal anti-PAR4 and monoclonal anti-PML antibodies, respectively. In another experiment, double staining revealed colocalization of Par4 and THAP1 in cells expressing ectopic GFP-THAP1. Primary human endothelial cells or fibroblasts were transfected with GFP-THAP1 expression vector; endogenous Par4 was stained with polyclonal anti-PAR4 antibodies.

The inventors further demonstrated that PML recruits the THAP1/Par4 complex to PML-NBs. Triple immunofluorescence staining showed colocalization of THAP1, Par4 and PML in cells overexpressing PML and absence of colocalization in cells expressing ectopic Sp100. Hela cells were cotransfected with GFP-THAP1 and HA-PML or HA-SP100 expression vectors; HA-PML or HA-SP100 and endogenous Par4 were stained with monoclonal anti-HA and polyclonal anti-Par4 antibodies, respectively.

**Assays to modulate THAP family protein trafficking in the PML nuclear bodies**

Provided are assays for the identification of drugs that modulate (stimulate or inhibit) THAP-family or THAP domain protein, particularly THAP1, binding to PML-NB proteins or localization to PML-NBs. In general, any suitable assay for the detection of protein-protein

interaction may be used. Two examples of high throughput screening assays include 1) a two hybrid-based assay in yeast to find compounds that disrupt interaction of the THAP1 bait with the PML-NB protein prey; and 2) in vitro interaction assays using recombinant THAP1 and PML-NB proteins. Such assays may be conducted as described above with respect to THAP-family/Par4 assays except that the PML-NB protein is used in place of Par4. Binding may be detected, for example, between a THAP-family protein and a PML protein or PML associated protein such as daxx, sp100, sp140, p53, pRB, CBP, BLM or SUMO-1.

Other assays for which standard methods are well known include assays to identify molecules that modulate, generally inhibit, the colocalization of THAP1 with PML-NBs.

Detection can be carried out using a suitable label, such as an anti-THAP1 antibody, and an antibody allowing the detection of PML-NB protein.

#### **Assays to modulate PAR4 trafficking in the PML bodies**

Provided are assays for the identification of drugs that modulate (stimulate or inhibit) PAR4 binding to PML-NB proteins or localization to PML-NBs. In general, any suitable assay for the detection of protein-protein interaction may be used. Two examples of high throughput screening assays include 1) a two hybrid-based assay in yeast to find compounds that disrupt interaction of the PAR4 bait with the PML-NB protein prey; and 2) in vitro interaction assays using recombinant PAR4 and PML-NB proteins. Such assays may be conducted as described above with respect to THAP-family polypeptide/Par4 assays except that the PML-NB protein is used in place of the THAP-family polypeptide. Binding may be detected, for example, between a Par4 protein and a PML protein or PML associated protein such as daxx, sp100, sp140, p53, pRB, CBP, BLM or SUMO-1.

Other assays for which standard methods are well known include assays to identify molecules that modulate, generally inhibit, the colocalization of PAR4 with PML-NBs. Detection can be carried out using a suitable label, such as an anti-PAR4 antibody, and an antibody allowing the detection of PML-NB protein.

This invention further pertains to novel agents identified by the above-described screening assays and to processes for producing such agents by use of these assays. Accordingly, in one embodiment, the present invention includes a compound or agent obtainable by a method comprising the steps of any one of the aforementioned screening assays (e.g., cell-based assays or cell-free assays). For example, in one embodiment, the invention includes a compound or agent obtainable by a method comprising contacting a cell which expresses a THAP-family target molecule with a test compound and determining the ability of the test compound to bind to, or modulate the activity of, the THAP-family target molecule. In another embodiment, the invention includes a compound or agent obtainable by a method comprising contacting a cell which expresses a THAP-family target molecule with a THAP-family protein or biologically-active portion thereof, to form an assay mixture, contacting the assay mixture with a test compound, and determining the

ability of the test compound to interact with, or modulate the activity of, the THAP-family target molecule. In another embodiment, the invention includes a compound or agent obtainable by a method comprising contacting a THAP-family protein or biologically active portion thereof with a test compound and determining the ability of the test compound to bind to, or modulate (e.g., stimulate or inhibit) the activity of, the THAP-family protein or biologically active portion thereof. In yet another embodiment, the present invention includes a compound or agent obtainable by a method comprising contacting a THAP-family protein or biologically active portion thereof with a known compound which binds the THAP-family protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with, or modulate the activity of the THAP-family protein.

Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a THAP-family or THAP domain modulating agent, an antisense THAP-family or THAP domain nucleic acid molecule, a THAP-family- or THAP domain- specific antibody, or a THAP-family- or THAP domain- binding partner) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

The present invention also pertains to uses of novel agents identified by the above-described screening assays for diagnoses, prognoses, and treatments as described herein. Accordingly, it is within the scope of the present invention to use such agents in the design, formulation, synthesis, manufacture, and/or production of a drug or pharmaceutical composition for use in diagnosis, prognosis, or treatment, as described herein. For example, in one embodiment, the present invention includes a method of synthesizing or producing a drug or pharmaceutical composition by reference to the structure and/or properties of a compound obtainable by one of the above-described screening assays. For example, a drug or pharmaceutical composition can be synthesized based on the structure and/or properties of a compound obtained by a method in which a cell which expresses a THAP-family target molecule is contacted with a test compound and the ability of the test compound to bind to, or modulate the activity of, the THAP-family target molecule is determined. In another exemplary embodiment, the present invention includes a method of synthesizing or producing a drug or pharmaceutical composition based on the structure and/or properties of a compound obtainable by a method in which a THAP-family protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to, or modulate (e.g., stimulate or inhibit) the activity of, the THAP-family protein or biologically active portion thereof is determined.

#### Apoptosis assays

It will be appreciated that any suitable apoptosis assay may be used to assess the apoptotic activity of a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof.

Apoptosis can be recognized by a characteristic pattern of morphological, biochemical, and molecular changes. Cells going through apoptosis appear shrunken, and rounded; they also can be observed to become detached from culture dish. The morphological changes involve a characteristic pattern of condensation of chromatin and cytoplasm which can be readily identified by microscopy. When stained with a DNA-binding dye, e.g., H33258, apoptotic cells display classic condensed and punctate nuclei instead of homogeneous and round nuclei.

A hallmark of apoptosis is endonucleolysis, a molecular change in which nuclear DNA is initially degraded at the linker sections of nucleosomes to give rise to fragments equivalent to single and multiple nucleosomes. When these DNA fragments are subjected to gel electrophoresis, they reveal a series of DNA bands which are positioned approximately equally distant from each other on the gel. The size difference between the two bands next to each other is about the length of one nucleosome, i.e., 120 base pairs. This characteristic display of the DNA bands is called a DNA ladder and it indicates apoptosis of the cell. Apoptotic cells can be identified by flow cytometric methods based on measurement of cellular DNA content, increased sensitivity of DNA to denaturation, or altered light scattering properties. These methods are well known in the art and are within the contemplation of the invention.

Abnormal DNA breaks which are characteristic of apoptosis can be detected by any means known in the art. In one preferred embodiment, DNA breaks are labeled with biotinylated dUTP (b-dUTP). As described in U.S. Patent No. 5,897,999, cells are fixed and incubated in the presence of biotinylated dUTP with either exogenous terminal transferase (terminal DNA transferase assay; TdT assay) or DNA polymerase (nick translation assay; NT assay). The biotinylated dUTP is incorporated into the chromosome at the places where abnormal DNA breaks are repaired, and are detected with fluorescein conjugated to avidin under fluorescence microscopy.

*Assessing THAP-family, THAPdomain and PAR4 polypeptides activity*

For assessing the nucleic acids and polypeptides of the invention, the apoptosis indicator which is assessed in the screening method of the invention may be substantially any indicator of the viability of the cell. By way of example, the viability indicator may be selected from the group consisting of cell number, cell refractility, cell fragility, cell size, number of cellular vacuoles, a stain which distinguishes live cells from dead cells, methylene blue staining, bud size, bud location, nuclear morphology, and nuclear staining. Other viability indicators and combinations of the viability indicators described herein are known in the art and may be used in the screening method of the invention.

Cell death status can be evaluated based on DNA integrity. Assays for this determination include assaying DNA on an agarose gel to identify DNA breaking into oligonucleosome ladders

and immunohistochemically detecting the nicked ends of DNA by labeling the free DNA end with fluorescein or horseradish peroxidase-conjugated UTP via terminal transferase. Routinely, one can also examine nuclear morphology by propidium iodide (PI) staining. All three assays (DNA ladder, end-labeling, and PI labelling) are gross measurements and good for those cells that are already  
5 dead or at the end stage of dying.

In a preferred example, an apoptosis assay is based on serum-withdrawal induced apoptosis in a 3T3 cell line with tetracycline-regulated expression of a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof. Detection of apoptotic cells is accomplished by TUNEL labeling cells in 96- or 384-well microplates. This example is further  
10 described in Example 23.

In other aspects, assays may test for the generation of cytotoxic death signals, anti-viral responses (Tartaglia et al., (1993) Cell 74(5):845-531), and/or the activation of acid sphingomyelinase (Wiegmann et al., (1994) Cell 78(6):1005-15) when the THAP-family protein is overexpressed or ectopically expressed in cells. Assaying for modulation of apoptosis can also be  
15 carried out in neuronal cells and lymphocytes for example, where factor withdrawal is known to induce cell suicide as demonstrated with neuronal cells requiring nerve growth factor to survive (Martin, D. P. et al, (1988) J. Cell Biol 106, 829-844) and lymphocytes depending on a specific lymphokine to live (Kyprianou, N. and Isaacs, J. T. (1988) Endocrinology 122:552-562).

*THAP-family or THAP domain polypeptide -marker fusions in cell assays*

In one method, an expression vector encoding the a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof can be used to evaluate the ability of the polypeptides of the invention to induce apoptosis in cells. If desired, a THAP-family or THAP domain polypeptide may be fused to a detectable marker in order to facilitate  
20 identification of those cells expressing the a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof. For example, a variant of the Aequoria victoria GFP variant, enhanced green fluorescent protein (EGFP), can be used in fusion protein production (CLONTECH Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, Calif. 94303), further described in U.S. Patent No. 6,191,269.

The THAP-family- or THAP domain polypeptide cDNA sequence is fused in-frame by  
30 insertion of the THAP-family- or THAP domain polypeptide encoding cDNA into the Sall-BamHI site of plasmid pEGFP-NI (GenBank Accession # U55762). Cells are transiently transfected by the method optimal for the cell being tested (either CaPO<sup>4</sup> or Lipofectin). Expression of a THAP-family or THAP domain polypeptide and induction of apoptosis is examined using a fluorescence microscope at 24 hrs and 48 hrs post-transfection. Apoptosis can be evaluated by the TUNEL  
35 method (which involves 3' end-labeling of cleaved nuclear and/or morphological criteria DNA) (Cohen et al. (1984) J. Immunol. 132:38-42). Where the screen uses a fusion polypeptide comprising a THAP-family or THAP domain polypeptide and a reporter polypeptide (e.g., EGFP),

apoptosis can be evaluated by detection of nuclear localization of the reporter polypeptide in fragmented nuclear bodies or apoptotic bodies. For example, where a THAP-family or THAP domain polypeptide-EGFP fusion polypeptide is used, distribution of THAP-family or THAP domain polypeptide EGFP-associated fluorescence in apoptotic cells would be identical to the distribution of DAPI or Hoechst 33342 dyes, which are conventionally used to detect the nuclear DNA changes associated with apoptosis (Cohen et al., supra). A minimum of approximately 100 cells, which display characteristic EGFP fluorescence, are evaluated by fluorescence microscopy. Apoptosis is scored as nuclear fragmentation, marked apoptotic bodies, and cytoplasmic boiling. The characteristics of nuclear fragmentation are particularly visible when THAP-family or THAP domain polypeptide-EGFP condenses in apoptotic bodies.

The ability of the THAP-family- or THAP domain polypeptides to undergo nuclear localization and to induce apoptosis can be tested by transient expression in 293 human kidney cells. If proved susceptible to THAP-family- or THAP domain- induced apoptosis, 293 cells can serve as a convenient initial screen for those THAP family or THAP domain polypeptides, or biologically active fragments or homologues thereof that will likely also induce apoptosis in other (e.g. endothelial cells or cancer cells). In an exemplary protocol, 293 cells are transfected with plasmid vectors expressing THAP-family- or THAP domain- EGFP fusion protein. Approximately  $5 \times 10^6$  293 cells in 100 mm dishes were transfected with 10 g of plasmid DNA using the calcium-phosphate method. The plasmids used are comprise CMV enhancer/promoter and THAP-family- or THAP domain- EGFP coding sequence). Apoptosis is evaluated 24 hrs after transfection by TUNEL and DAPI staining. The THAP-family- or THAP domain- EGFP vector transfected cells are evaluated by fluorescence microscopy with observation of typical nuclear aggregation of the EGFP marker as an indication of apoptosis. If apoptotic, the distribution of EGFP signal in cells expressing THAP-family- or THAP domain-EGFP will be identical to the distribution of DAPI or Hoechst 33342 dyes, which are conventionally used to detect the nuclear DNA changes associated with apoptosis (Cohen et al., supra).

The ability of the THAP family or THAP domain polypeptides, or biologically active fragments or homologues thereof to induce apoptosis can also be tested by expression assays in human cancer cells, for example as available from NCI. Vector type (for example plasmid or retroviral or sindbis viral) can be selected based on efficiency in a given cell type. After the period indicated, cells are evaluated for morphological signs of apoptosis, including aggregation of THAP-family- or THAP domain- EGFP into nuclear apoptotic bodies. Cells are counted under a fluorescence microscope and scored as to the presence or absence of apoptotic signs, or cells are scored by fluorescent TUNEL assay and counted in a flow cytometer. Apoptosis is expressed as a percent of cells displaying typical advanced changes of apoptosis.

Cells from the NCI panel of tumor cells include from example:

-colon cancer, expression using a retroviral expression vector, with evaluation of apoptosis at 96 hrs post-infection (cell lines KM12; HT-29; SW-620; COLO205; HCT-5; HCC 2998; HCT-116);

5 -CNS tumors, expression using a retroviral expression vector, with evaluation of apoptosis at 96 hrs post-infection (cell lines SF-268, astrocytoma; SF-539, glioblastoma; SNB-19, glioblastoma; SNB-75, astrocytoma; and U251, glioblastoma;

-leukemia cells, expression using a retroviral expression vector, with evaluation of apoptosis at 96 hrs post-infection (cell lines CCRF-CEM, acute lymphocytic leukemia (ALL); K562, acute myelogenous leukemia (AML); MOLT-4, ALL; SR, immunoblastoma large cell; and  
10 RPMI 8226, Myeloblastoma);

-prostate cancer, expression using a retroviral expression vector, with evaluation of apoptosis at 96 hrs post-infection (PC-3);

-kidney cancer, expression using a retroviral expression vector, with evaluation of apoptosis at 96 hrs post-infection (cell lines 768-0; UO-31; TK10; ACHN);

15 -skin cancer, expression using a retroviral expression vector, with evaluation of apoptosis at 96 hrs post-infection (Melanoma) (cell lines SKMEL-28; M14; SKMEL-5; MALME-3);

-lung cancer, expression using a retroviral expression vector, with evaluation of apoptosis at 96 hrs post-infection (cell lines HOP-92; NCI-H460; HOP-62; NCI-H522; NCI-H23; A549; NCI-H226; EKVX; NCI-H322);

20 -breast cancer, expression using a retroviral expression vector, with evaluation of apoptosis at 96 hrs post-infection (cell lines MCF-7; T-47D; MCF-7/ADR; MDAMB43; MDAMB23; MDA-N; BT-549);

-ovary cancer, expression using either a retroviral expression vector and protocol or the Sindbis viral expression vector and protocol, with evaluation of apoptosis at 96 hrs post-infection with retrovirus or at 24 hrs post-infection with Sindbis viral vectors (cell lines OVCAR-8; OVCAR-4; IGROV-1; OVCAR-5; OVCAR3; SK-OV-3).  
25

In a further representative example, the susceptibility of malignant melanoma cells to apoptosis induced by a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof can be tested in several known melanoma cell types: human  
30 melanoma WM 266-4 (ATCC CRL-1676); human malignant melanoma A-375 (ATCC CRL-1619); human malignant, melanoma A2058 (ATCC CRL-11147); human malignant melanoma SK-MEL-31 (ATCC HTB-73); human malignant melanoma RPMI-7591 ATCC HTB-66 (metastasis to lymph node). Primary melanoma isolates can also be tested. In addition, human chronic myelogenous leukemia K-562 cells (ATCC CCL-243), and 293 human kidney cells (ATCC CRL-  
35 1573) (transformed primary embryonal cell) are tested. Normal human primary dermal fibroblasts and Rat-1 fibroblasts serve as controls. All melanoma cell lines are metastatic on the basis of their isolation from metastases or metastatic nodules. A transient expression strategy is used in order to

**WO 2004/055050**

**PCT/IB2003/006434**



cells) are plated in 24 well dishes at  $3.5 \times 10^4$  cells/well. The following day, the cells are transfected with a marker plasmid encoding beta-galactosidase, in combination with an expression plasmid encoding THAP-family or THAP domain polypeptide, by the Lipofectamine procedure (Gibco/BRL). At 24 hours post transfection, cells are fixed and stained with X-Gal to detect beta-galactosidase expression in cells that received plasmid DNA (Miura et al., supra). The number of blue cells is counted by microscopic examination and scored as either live (flat blue cells) or dead (round blue cells). The cell killing activity of the THAP-family or THAP domain polypeptide in this assay is manifested by a large reduction in the number of blue cells obtained relative to co-transfection of the beta-gal plasmid with a control expression vector (i.e., with no THAP-family or THAP domain polypeptide cDNA insert).

In yet another example, beta-galactosidase co-transfection assays can be used for determination of cell death. The assay is performed as described (Hsu, H. et al, (1995). Cell 81,495-504; Hsu, H. et al, (1996a). Cell 84, 299-308; and Hsu, H. et al, (1996b) Immunity 4, 387-396 and U.S. Patent No. 6,242,569). Transfected cells are stained with X-gal as described in Shu, H. B. et al, ((1995) J. Cell Sci. 108, 2955-2962). The number of blue cells from 8 viewing fields of a 35 mm dish is determined by counting. The average number from one representative experiment is shown.

Assays for apoptosis can also be carried out by making use of any suitable biological marker of apoptosis. Several methods are described as follows.

In one aspect, fluorocytometric studies of cell death status can be carried out. Technology used in fluorocytometric studies employs the identification of cells at three different phases of the cell cycle: G<sub>1</sub>, S, and G<sub>2</sub>. This is largely performed by DNA quantity staining by propidium iodide labeling. Since the dying cell population contains the same DNA quantity as the living counterparts at any of the three phases of the cell cycle, there is no way to distinguish the two cell populations. One can perform double labeling for a biological marker of apoptosis (e.g. terminin Tp30, U.S. Patent No. 5,783,667) positivity and propidium iodide (PI) staining together. Measurement of the labeling indices for the biological marker of apoptosis and PI staining can be used in combination to obtain the exact fractions of those cells in G<sub>1</sub> that are living and dying. Similar estimations can be made for the S-phase and G<sub>2</sub> phase cell populations.

In this assay, the cells are processed for formaldehyde fixation and extraction with 0.05% Triton. Afterwards, the cell specimens are incubated with monoclonal antibody to a marker of apoptosis overnight at room temperature or at 37C for one hour. This is followed by further incubation with fluoresceinated goat antimouse antibody, and subsequent incubation by propidium iodide staining. The completely processed cell specimens are then evaluated by fluorocytometric measurement on both fluorescence (marker of apoptosis) and rhodamine (PI) labeling intensity on a per cell basis, with the same cell population simultaneously.

In another aspect, it is possible to assess the inhibitory effect on cell growth by therapeutic induction of apoptosis. One routine method to determine whether a particular chemotherapeutic

drug can inhibit cancerous cell growth is to examine cell population size either in culture, by measuring the reduction in cell colony size or number, or measuring soft agar colony growth or in vivo tumor formation in nude mice, which procedures require time for development of the colonies or tumor to be large enough to be detectable. Experiments involved in these approaches in general require large-scale planning and multiple repeats of lengthy experimental span (at least three weeks). Often these assays do not take into account the fact that a drug may not be inhibiting cell growth, but rather killing the cells, a more favorable consequence needed for chemotherapeutic treatment of cancer. Thus, assays for the assessment of apoptosis activity can involve using a biological or biochemical marker specific for quiescent, non-cycling or non-proliferating cells. For example, a monoclonal antibody can be used to assess the non-proliferating population of cells in a given tissue which indirectly gives a measure of the proliferating component of a tumor or cell mass. This detection can be combined with a biological or biochemical marker (e.g. antibodies) to detect the dying cell population pool, providing a powerful and rapid assessment of the effectiveness of any given drugs in the containment of cancerous cell growth. Applications can be easily performed at the immunofluorescence microscopic level with cultured cells or tissue sections.

In other aspects, a biological or biochemical marker can be used to assess pharmacological intervention on inhibition of cell death frequency in degenerative diseases. For degenerative diseases such as Alzheimer's or Parkinson's disease, these losses may be due to the premature activation of the cell death program in neurons. In osteoporosis, the cell loss may be due to an improper balance between osteoblast and osteoclast cells, due to the too active programmed cell death process killing more cells than the bone tissue can afford. Other related phenomena may also occur in the wound healing process, tissue transplantation and cell growth in the glomerus during kidney infection, where the balance between living and dying cell populations is an essential issue to the health status of the tissue, and are further described in the section titled "Methods of treatment". A rapid assessment of dying cell populations can be made through the immunohistochemical and biochemical measurements of a biological or biochemical marker of apoptosis in degenerative tissues. In one example, a biological or biochemical marker can be used to assess cell death status in oligodendrocytes associated with Multiple Sclerosis. Positive staining of monoclonal antibody to a marker of apoptosis (such as Tp30, U.S. Patent No. 5,783,667) occurs in dying cultured human oligodendrocytes. The programmed cell death event is activated in these oligodendrocytes by total deprivation of serum, or by treatment with tumor necrosis factor (TNF).

In general, a biological or biochemical marker can also be used to assess cell death status in pharmacological studies in animal models. Attempting to control either a reduced cell death rate, in the case of cancer, or an increased cell death rate, in the case of neurodegeneration, has been recently seen as a new mode of disease intervention. Numerous approaches via either intervention with known drugs or gene therapy are in progress, starting from the base of correcting the altered

programmed cell death process, with the concept on maintaining a balanced cell mass in any given tissue. For these therapeutic interventions, the bridge between studies in cultured cells and clinical trials is animal studies, i.e. success in intervention with animal models, in either routine laboratory animals or transgenic mice bearing either knock-out or overexpression phenotypes. Thus, a biological or biochemical marker of apoptosis, such as an antibody for an apoptosis-specific protein, is a useful tool for examining apoptotic death status in terms of change in dying cell numbers between normal and experimentally manipulated animals. In this context the invention, as a diagnostic tool for assessing cell death status, could help to determine the efficacy and potency of a drug or a gene therapeutic approach.

As discussed, provided are methods for assessing the activity of THAP-family members and therapeutic treatment acting on THAP-family members or related biological pathways. However, in other aspects, the same methods may be used for assessment of apoptosis in general, when a THAP-family member is used as a biological marker of apoptosis. Thus, the invention also provides diagnostic and assay methods using a THAP-family member as a marker of cell death or apoptotic activity. Further diagnostic assays are also provided herein in the section titled 'Diagnostic and prognostic uses'.

#### METHODS OF TREATMENT

A large body of evidence gathered from experiments carried out with apoptosis modulating strategies suggests that treatments acting on apoptosis-inducing or cell proliferation-reducing proteins may offer new treatment methods for a wide range of disorders. Methods of treatment according to the invention may act in a variety of manners, given the novel function provided for a number of proteins, and the linking of several biological pathways.

Provided herein are treatment methods based on the functionalization of the THAP-family members. THAP family or THAP domain polypeptides, and biologically active fragments and homologues thereof, as described further herein may be useful in modulation of apoptosis or cell proliferation.

The methods of treatment involve acting on a molecule of the invention (that is, a THAP family member polypeptide, THAP-family target, or PAR4 or PAR4 target). Included are methods which involve modulating THAP-family polypeptide activity, THAP-family target activity, or PAR4 or PAR4 target activity. This modulation (increasing or decreasing) of activity can be carried out in a number of suitable ways, several of which have been described in the present application.

For example, methods of treatment may involve modulating a "THAP-family activity", "biological activity of a THAP-family member" or "functional activity of a THAP-family member". Modulating THAP-family activity may involve modulating an association with a THAP-family-target molecule (for example, association of THAP1, THAP2 or THAP3 with Par4 or association of THAP1, THAP2 or THAP3 with a PML-NB protein) or preferably any other activity selected from

the group consisting of: (1) mediating apoptosis or cell proliferation when expressed or introduced into a cell, most preferably inducing or enhancing apoptosis, and/or most preferably reducing cell proliferation; (2) mediating apoptosis or cell proliferation of an endothelial cell; (3) mediating apoptosis or cell proliferation of a hyperproliferative cell; (4) mediating apoptosis or cell proliferation of a CNS cell, preferably a neuronal or glial cell; or (5) an activity determined in an animal selected from the group consisting of mediating, preferably inhibiting angiogenesis, mediating, preferably inhibiting inflammation, inhibition of metastatic potential of cancerous tissue, reduction of tumor burden, increase in sensitivity to chemotherapy or radiotherapy, killing a cancer cell, inhibition of the growth of a cancer cell, or induction of tumor regression. Detecting THAP-family activity may also comprise detecting any suitable therapeutic endpoint associated with a disease condition discussed herein.

In another example, methods of treatment may involve modulating a "PAR4 activity", "biological activity of PAR4" or "functional activity of PAR4 ". Modulating PAR4 activity may involve modulating an association with a PAR4-target molecule (for example THAP1, THAP2, THAP3 or PML-NB protein) or most preferably PAR4 apoptosis inducing or enhancing (e.g. signal transducing) activity, or inhibition of cell proliferation or cell cycle.

Methods of treatment may involve modulating the recruitment, binding or association of proteins to PML-NBs, or otherwise modulating PML-NBs activity. The present invention also provides methods for modulating PAR4 activity, comprising modulating PAR4 interactions with THAP-family proteins, and PAR4 and PML-NBs, as well as modulating THAP-family activity, comprising modulating for example THAP1 interactions with PML-NBs. The invention encompasses inhibiting or increasing the recruitment of THAP1, or PAR4 to PML-NBs. Preventing the binding of either or both of THAP1 or PAR4 to PML-NBs may increase the bioavailability of THAP1 and/or PAR4, thus providing a method of increasing THAP1 and/or PAR4 activity. The invention also encompasses inhibiting or increasing the binding of a THAP-family protein (such as THAP1) or PAR4 to PML-NBs or to another protein associated with PML-NBs, such as a protein selected from the group consisting of daxx, sp100, sp140, p53, pRB, CBP, BLM, SUMO-1. For example, the invention encompasses modulating PAR4 activity by preventing the binding of THAP1 to PAR4, or by preventing the recruitment or binding of PAR4 to PML-NBs.

Therapeutic methods and compositions of the invention may involve (1) modulating apoptosis or cell proliferation, most preferably inducing or enhancing apoptosis, and/or most preferably reducing cell proliferation; (2) modulating apoptosis or cell proliferation of an endothelial cell (3) modulating apoptosis or cell proliferation of a hyperproliferative cell; (4) modulating apoptosis or cell proliferation of a CNS cell, preferably a neuronal or glial cell; (5) inhibition of metastatic potential of cancerous tissue, reduction of tumor burden, increase in sensitivity to chemotherapy or radiotherapy, killing a cancer cell, inhibition of the growth of a cancer cell, or induction tumor regression; or (6) interaction with a THAP family target molecule or

THAP domain target molecule, preferably interaction with a protein or a nucleic acid. Methods may also involve improving a symptom of or ameliorating a condition as further described herein.

#### *Antiapoptotic therapy*

Molecules of the invention (e.g. those obtained using the screening methods described herein, dominant negative mutants, antibodies etc.) which inhibit apoptosis are also expected to be useful in the treatment and/or prevention of disease. Diseases in which it is desirable to prevent apoptosis include neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, retinitis pigmentosa and cerebellar degeneration; myelodysplasia such as aplastic anemia; ischemic diseases such as myocardial infarction and stroke; hepatic diseases such as alcoholic hepatitis, hepatitis B and hepatitis C; joint-diseases such as osteoarthritis; atherosclerosis; and etc. The apoptosis inhibitor of the present invention is especially preferably used as an agent for prophylaxis or treatment of a neurodegenerative disease (see also Adams, J. M., Science, 281:1322 (1998).

Included as inhibitors of apoptosis as described herein are generally any molecule which inhibits activity of a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof, a THAP-family target protein or PAR4 (particularly PAR4/PML-NB protein interactions). THAP-family and THAP domain polypeptides inhibitors may include for example antibodies, peptides, dominant negative THAP-family or THAP domain analogs, small molecules, ribozyme or antisense nucleic acids. These inhibitors may be particularly advantageous in the treatment of neurodegenerative disorders. Particularly preferred are inhibitors which affect binding of THAP-family protein to a THAP-family target protein, and inhibitors which affect the DNA binding activity of a THAP-family protein.

In further preferred aspects the invention provides inhibitors of THAP-family activity, including but not limited to molecules which interfere or inhibit interactions of THAP-family proteins with PAR4, for the treatment of endothelial cell related disorders and neurodegenerative disorders. Support is found in the literature, as PAR4 appears to play a key role in neuronal apoptosis in various neurodegenerative disorders (Guo et al., 1998; Mattson et al., 2000; Mattson et al., 1999; Mattson et al., 2001). THAP1, which is expressed in brain and associates with PAR4 may therefore also play a key role in neuronal apoptosis. Drugs that inhibit THAP-family and/or inhibit THAP-family/PAR4 complex formation may lead to the development of novel preventative and therapeutic strategies for neurodegenerative disorders.

#### *Apoptosis regulation in endothelial cells*

The invention also provides methods of regulating angiogenesis in a subject which are expected to be useful in the treatment of cancer, cardiovascular diseases and inflammatory diseases. An inducer of apoptosis of immortalized cells is expected to be useful in suppressing tumorigenesis and/or metastasis in malignant tumors. Examples of malignant tumors include leukemia (for example, myelocytic leukemia, lymphocytic leukemia such as Burkitt lymphoma), digestive tract

carcinoma, lung carcinoma, pancreas carcinoma, ovary carcinoma, uterus carcinoma, brain tumor, malignant melanoma, other carcinomas, and sarcomas. The present inventors have isolated both THAP1 and PAR4 cDNAs from human endothelial cells, and both PAR4 and PML are known to be expressed predominantly in blood vessel endothelial cells (Boghaert et al., (1997) Cell Growth Differ 8(8):881-90; Terris B. et al, (1995) Cancer Res. 55(7):1590-7, 1995), suggesting that the PML-NBs and the newly associated THAP1/PAR4 proapoptotic complex may be a major regulator of endothelial cell apoptosis *in vivo* and thus constitute an attractive therapeutic target for angiogenesis-dependent diseases. For example, THAP1 and PAR4 pathways may allow selective treatments that regulate (e.g. stimulate or inhibit) angiogenesis.

10 In a first aspect, the invention provides methods of inhibiting endothelial cell apoptosis, by administering a THAP1 or PAR4 inhibitor, or optionally a THAP1/PAR4 interaction inhibitor or optionally an inhibitor of THAP1 DNA binding activity. As further described herein, the THAP domain is involved in THAP1 pro-apoptotic activity. Deletion of the THAP domain abrogates the proapoptotic activity of THAP1 in mouse 3T3 fibroblasts, as shown in Example 11. Also, as  
15 further described herein, deletion of residues 168-172 or replacement of residues 171-172 abrogates THAP1 binding to PAR4 both *in vitro* and *in vivo* and results in lack of recruitment of PAR4 by THAP1 to PML-NBs. For PAR4, the leucine zipper domain is required (and is sufficient) for binding to THAP1.

Inhibiting endothelial cell apoptosis may improve angiogenesis and vasculogenesis in  
20 patients with ischemia and may also interfere with focal dysregulated vascular remodeling, the key mechanism for atherosclerotic disease progression.

In another aspect, the invention provides methods of inducing endothelial cell apoptosis, by administering for example a biologically active THAP family polypeptide such as THAP1, a THAP domain polypeptide or a PAR4 polypeptide, or a biologically active fragment or homologue  
25 thereof, or a THAP1 or PAR4 stimulator. Stimulation of endothelial cell apoptosis may prevent or inhibit angiogenesis and thus limit unwanted neovascularization of tumors or inflamed tissues (see Dimmeler and Zeiher, Circulation Research, 2000, 87 :434-439).

#### *Angiogenesis*

Angiogenesis is defined in adult organism as the formation of new blood vessels by a  
30 process of sprouting from pre-existing vessels. This neovascularization involves activation, migration, and proliferation of endothelial cells and is driven by several stimuli, among those shear stress. Under normal physiological conditions, humans or animals undergo angiogenesis only in very specific restricted situations. For example, angiogenesis is normally observed in wound healing, fetal and embryonal development and formation of the corpus luteum, endometrium and  
35 placenta. Molecules of the invention may have endothelial inhibiting or inducing activity, having the capability to inhibit or induce angiogenesis in general.

Both controlled and uncontrolled angiogenesis are thought to proceed in a similar manner. Endothelial cells and pericytes, surrounded by a basement membrane, form capillary blood vessels. Angiogenesis begins with the erosion of the basement membrane by enzymes released by endothelial cells and leukocytes. The endothelial cells, which line the lumen of blood vessels, then protrude through the basement membrane. Angiogenic stimulants induce the endothelial cells to migrate through the eroded basement membrane. The migrating cells form a "sprout" off the parent blood vessel, where the endothelial cells undergo mitosis and proliferate. The endothelial sprouts merge with each other to form capillary loops, creating the new blood vessel.

Persistent, unregulated angiogenesis occurs in a multiplicity of disease states, tumor metastasis and abnormal growth by endothelial cells and supports the pathological damage seen in these conditions. The diverse pathological disease states in which unregulated angiogenesis is present have been grouped together as angiogenic dependent or angiogenic associated diseases. It is thus an object of the present invention to provide methods and compositions for treating diseases and processes that are mediated by angiogenesis including, but not limited to, hemangioma, solid tumors, leukemia, metastasis, telangiectasia psoriasis scleroderma, pyogenic granuloma, Myocardial angiogenesis, plaque neovascularization, coronary collaterals, ischemic limb angiogenesis, corneal diseases, rubeosis, neovascular glaucoma, diabetic retinopathy, retrolental fibroplasia, arthritis, diabetic neovascularization, macular degeneration, wound healing, peptic ulcer, fractures, keloids, vasculogenesis, hematopoiesis, ovulation, menstruation, and placentation.

*(i) Anti-angiogenic therapy*

In one aspect the invention provides anti-angiogenic therapies as potential treatments for a wide variety of diseases, including cancer, arteriosclerosis, obesity, arthritis, duodenal ulcers, psoriasis, proliferative skin disorders, cardiovascular disorders and abnormal ocular neovascularization caused, for example, by diabetes (Folkman, Nature Medicine 1:27 (1995) and Folkman, Seminars in Medicine of the Beth Israel Hospital, Boston, New England Journal of Medicine, 333:1757 (1995)). Anti-angiogenic therapies are thought to act by inhibiting the formation of new blood vessels.

The present invention thus provides methods and compositions for treating diseases and processes mediated by undesired and uncontrolled angiogenesis by administering to a human or animal a composition comprising a substantially purified THAP family or THAP domain polypeptide, or a biologically active fragment, homologue or derivative thereof in a dosage sufficient to inhibit angiogenesis, administering a vector capable of expressing a nucleic acid encoding a THAP-family or THAP domain protein, or administering any other inducer of expression or activity of a THAP-family or THAP domain protein. The present invention is particularly useful for treating or for repressing the growth of tumors. Administration of THAP-family or THAP domain nucleic acid, protein or other inducer to a human or animal with prevascularized metastasized tumors will prevent the growth or expansion of those tumors. THAP-

family activity may be used in combination with other compositions and procedures for the treatment of diseases. For example, a tumor may be treated conventionally with surgery, radiation or chemotherapy combined with THAP-family or THAP domain protein and then THAP-family or THAP domain protein may be subsequently administered to the patient to extend the dormancy of micrometastases and to stabilize any residual primary tumor.

In a preferred example, a THAP-family polypeptide activity, preferably a THAP1 activity is used for the treatment of arthritis, for example rheumatoid arthritis. Rheumatoid arthritis is characterized by symmetric, polyarticular inflammation of synovial-lined joints, and may involve extraarticular tissues, such as the pericardium, lung, and blood vessels.

(ii) *Angiogenic therapy*

In another aspect, the inhibitors of THAP-family protein activity, particularly THAP1 activity, could be used as an anti-apoptotic and thus as an angiogenic therapy. Angiogenic therapies are potential treatments for promoting wound healing and for stimulating the growth of new blood vessels to by-pass occluded ones. Thus, pro-angiogenic therapies could potentially augment or replace by-pass surgeries and balloon angioplasty (PTCA). For example, with respect to neovascularization to bypass occluded blood vessels, a "therapeutically effective amount" is a quantity which results in the formation of new blood vessels which can transport at least some of the blood which normally would pass through the blocked vessel.

The THAP-family protein of the present invention can for example be used to generate antibodies that can be used as inhibitors of apoptosis. The antibodies can be either polyclonal antibodies or monoclonal antibodies. In addition, these antibodies that specifically bind to the THAP-family protein can be used in diagnostic methods and kits that are well known to those of ordinary skill in the art to detect or quantify the THAP-family protein in a body fluid. Results from these tests can be used to diagnose or predict the occurrence or recurrence of a cancer and other angiogenic mediated diseases.

It will be appreciated that other inhibitors of THAP-family and THAP domain proteins can also be used in angiogenic therapies, including for example small molecules, antisense nucleic acids, dominant negative THAP-family and THAP domain proteins or peptides identified using the above methods.

In view of applications in both angiogenic and antiangiogenic therapies, molecules of the invention may have endothelial inhibiting or inducing activity, having the capability to inhibit or induce angiogenesis in general. It will be appreciated that methods of assessing such capability are known in the art, including for example assessing antiangiogenic properties as the ability inhibit the growth of bovine capillary endothelial cells in culture in the presence of fibroblast growth factor.

It is to be understood that the present invention is contemplated to include any derivatives of the THAP family or THAP domain polypeptides, and biologically active fragments and homologues thereof that have endothelial inhibitory or apoptotic activity. The present invention



includes full-length THAP-family and THAP domain proteins, derivatives of the THAP-family and THAP domain proteins and biologically-active fragments of the THAP-family and THAP domain proteins. These include proteins with THAP-family protein activity that have amino acid substitutions or have sugars or other molecules attached to amino acid functional groups. The methods also contemplate the use of genes that code for a THAP-family protein and to proteins that are expressed by those genes.

As discussed, several methods are described herein for delivering a modulator to a subject in need of treatment, including for example small molecule modulators, nucleic acids including via gene therapy vectors, and polypeptides including peptide mimetics, active polypeptides, dominant negative polypeptides and antibodies. It will be thus be appreciated that modulators of the invention identified according to the methods in the section titled "Drug Screening Assays" can be further tested in cell or animal models for their ability to ameliorate or prevent a condition involving a THAP-family polypeptide, particularly THAP1, THAP1, THAP2 or THAP3/PAR4 interactions, THAP-family DNA binding or PAR4 / PML-NBs interactions. Likewise, nucleic acids, polypeptides and vectors (e.g. viral) can also be assessed in a similar manner.

An "individual" treated by the methods of this invention is a vertebrate, particularly a mammal (including model animals of human disease, farm animals, sport animals, and pets), and typically a human.

"Treatment" refers to clinical intervention in an attempt to alter the natural course of the individual being treated, and may be performed either for prophylaxis or during the course of clinical pathology. Desirable effects include preventing occurrence or recurrence of disease, alleviation of symptoms, diminishment of any direct or indirect pathological consequences of the disease, such as hyperresponsiveness, inflammation, or necrosis, lowering the rate of disease progression, amelioration or palliation of the disease state, and remission or improved prognosis. The "pathology" associated with a disease condition is anything that compromises the well-being, normal physiology, or quality of life of the affected individual.

Treatment is performed by administering an effective amount of a THAP-family polypeptide inhibitor or activator. An "effective amount" is an amount sufficient to effect a beneficial or desired clinical result, and can be administered in one or more doses.

The criteria for assessing response to therapeutic modalities employing the lipid compositions of this invention are dictated by the specific condition, measured according to standard medical procedures appropriate for the condition.

#### **REDUCING CHEMOKINE MEDIATED EFFECTS**

Some aspects of the present invention relate to the use of THAP-family polypeptides, including THAP-1, chemokine-binding domains of THAP-family polypeptides, THAP-family polypeptide or THAP-family chemokine-binding domain fusions to immunoglobulin Fc, oligomers of THAP-family polypeptides or THAP-family chemokine-binding domains, or homologs of any of

the above-listed compositions (together and herein after referred to as THAP-type chemokine-binding agents) for reducing the inflammation or the symptoms associated with diseases or conditions that are influenced or mediated by chemokine binding or activity. In such embodiments, the THAP-type chemokine binding agents are administered to a subject in effective amounts so as to reduce the symptoms associated with the condition. In some embodiments, the chemokine that is effected by the THAP-type chemokine binding agent is SLC, CCL19, CCL5, CXCL9, CXCL10 or a combination of these chemokines. In other embodiments, the chemokine that is effected by the THAP-type chemokine binding agent is XCL1, XCL2, CCL1, CCL2, CCL3, CCL3L1, SCYA3L2, CCL4, CCL4L, CCL5, CCL6, CCL7, CCL8, SCYA9, SCYA10, CCL11, SCYA12, CCL13, CCL14, CCL15, CCL16, CCL17, CCL18, CCL19, CCL20, CCL21, CCL22, CCL23, CCL24, CCL25, CCL26, CCL27, CCL28, clone 391, CARP CC-1, CCL1, CK-1, regakine-1, K203, CXCL1, CXCL1P, CXCL2, CXCL3, PF4, PF4V1, CXCL5, CXCL6, PPBP, SPBPBP, IL8, CXCL9, CXCL10, CXCL11, CXCL12, CXCL14, CXCL15, CXCL16, NAP-4, LFCA-1, Scyba, JSC, VHSV-induced protein, CX3CL1, fCL1 or a combination of these chemokines. In some embodiments, the THAP-type chemokine-binding agent is administered directly whereas in other embodiments it is administered as a pharmaceutical composition. In either case, the routes of administration that are known in the art and described herein may be used to deliver the THAP-type chemokine-binding agent to the subject.

Some embodiments of the present invention relate to a device for delivering the THAP-type chemokine-binding agent or pharmaceutical composition thereof to the subject. In such embodiment, the device comprises a container which contains the THAP-type chemokine-binding agent or pharmaceutical composition thereof. For example, in some embodiments, the device may be a conventional device including, but not limited to, syringes, devices for intranasal administration of compositions and vaccine guns. In one embodiment, the device comprises a member which receives the THAP-type chemokine-binding agent or pharmaceutical composition thereof in communication with a mechanism for delivering the composition to the subject. In some embodiments, the device is an inhaler or a patch for transdermal administration.

#### **Pharmaceutical Compositions**

Compounds capable of inhibiting THAP-family activity, preferably small molecules but also including peptides, THAP-family nucleic acid molecules, THAP-family proteins, and anti-THAP-family antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional

media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

5 A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; 10 buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

15 Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL $\alpha$  (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent 20 that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use 25 of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which 30 delays absorption, for example, aluminum monostearate and gelatin.

Where the active compound is a protein, peptide or anti-THAP-family antibody, sterile injectable solutions can be prepared by incorporating the active compound (e.g., ) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as 35 required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of

sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer. Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art. Most preferably, active compound is delivered to a subject by intravenous injection.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Pat. No. 4,522,811.

It is especially advantageous to formulate oral or preferably parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD50

(the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to  
5 design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The  
10 dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as  
15 determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

20 It will be appreciated that THAP-type chemokine-binding agents can be formulated as pharmaceutical compositions and administered as described above. Additionally, the effective dose, route of administration, duration of administration, duration between doses and therapeutic effect can be determined by the methods described above as well as using methods that are well known in the art.

## 25 **Diagnostic and Prognostic Uses**

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics; and in drug screening and methods of treatment (e.g., therapeutic and prophylactic) as further described herein.

30 The invention provides diagnostic and prognostic assays for detecting THAP-family members, as further described. Also provided are diagnostic and prognostic assays for detecting interactions between THAP-family members and THAP-family target molecules. In a preferred example, a THAP-family member is THAP1, THAP2 or THAP3 and the THAP-family target is PAR4 or a PML-NB protein.

35 The invention also provides diagnostic and prognostic assays for detecting THAP1 and/or PAR4 localization to or association with PML-NBs, or association with or binding to a PML-NB-associated protein, such as daxx, sp100, sp140, p53, pRB, CBP, BLM or SUMO-1. In a preferred

method, the invention provides detecting PAR4 localization to or association with PML-NBs. In a further aspect, the invention provides detecting THAP-family nucleic acid binding activity.

The isolated nucleic acid molecules of the invention can be used, for example, to detect THAP-family polypeptide mRNA (e.g., in a biological sample) or a genetic alteration in a THAP-family gene, and to modulate a THAP-family polypeptide activity, as described further below. The THAP-family proteins can be used to treat disorders characterized by insufficient or excessive production of a THAP-family protein or THAP-family target molecules. In addition, the THAP-family proteins can be used to screen for naturally occurring THAP-family target molecules, to screen for drugs or compounds which modulate, preferably inhibit THAP-family activity, as well as to treat disorders characterized by insufficient or excessive production of THAP-family protein or production of THAP-family protein forms which have decreased or aberrant activity compared to THAP-family wild type protein. Moreover, the anti-THAP-family antibodies of the invention can be used to detect and isolate THAP-family proteins, regulate the bioavailability of THAP-family proteins, and modulate THAP-family activity.

Accordingly one embodiment of the present invention involves a method of use (e.g., a diagnostic assay, prognostic assay, or a prophylactic/therapeutic method of treatment) wherein a molecule of the present invention (e.g., a THAP-family protein, THAP-family nucleic acid, or most preferably a THAP-family inhibitor or activator) is used, for example, to diagnose, prognose and/or treat a disease and/or condition in which any of the aforementioned THAP-family activities is indicated. In another embodiment, the present invention involves a method of use (e.g., a diagnostic assay, prognostic assay, or a prophylactic/therapeutic method of treatment) wherein a molecule of the present invention (e.g., a THAP-family protein, THAP-family nucleic acid, or a THAP-family inhibitor or activator) is used, for example, for the diagnosis, prognosis, and/or treatment of subjects, preferably a human subject, in which any of the aforementioned activities is pathologically perturbed. In a preferred embodiment, the methods of use (e.g., diagnostic assays, prognostic assays, or prophylactic/therapeutic methods of treatment) involve administering to a subject, preferably a human subject, a molecule of the present invention (e.g., a THAP-family protein, THAP-family nucleic acid, or a THAP-family inhibitor or activator) for the diagnosis, prognosis, and/or therapeutic treatment. In another embodiment, the methods of use (e.g., diagnostic assays, prognostic assays, or prophylactic/therapeutic methods of treatment) involve administering to a human subject a molecule of the present invention (e.g., a THAP-family protein, THAP-family nucleic acid, or a THAP-family inhibitor or activator).

For example, the invention encompasses a method of determining whether a THAP-family member is expressed within a biological sample comprising: a) contacting said biological sample with: ii) a polynucleotide that hybridizes under stringent conditions to a THAP-family nucleic acid; or iii) a detectable polypeptide (e.g. antibody) that selectively binds to a THAP-family polypeptide; and b) detecting the presence or absence of hybridization between said polynucleotide and an RNA

species within said sample, or the presence or absence of binding of said detectable polypeptide to a polypeptide within said sample. A detection of said hybridization or of said binding indicates that said THAP-family member is expressed within said sample. Preferably, the polynucleotide is a primer, and wherein said hybridization is detected by detecting the presence of an amplification product comprising said primer sequence, or the detectable polypeptide is an antibody.

Also envisioned is a method of determining whether a mammal, preferably human, has an elevated or reduced level of expression of a THAP-family member, comprising: a) providing a biological sample from said mammal; and b) comparing the amount of a THAP-family polypeptide or of a THAP-family RNA species encoding a THAP-family polypeptide within said biological sample with a level detected in or expected from a control sample. An increased amount of said THAP-family polypeptide or said THAP-family RNA species within said biological sample compared to said level detected in or expected from said control sample indicates that said mammal has an elevated level of THAP-family expression, and a decreased amount of said THAP-family polypeptide or said THAP-family RNA species within said biological sample compared to said level detected in or expected from said control sample indicates that said mammal has a reduced level of expression of a THAP-family member.

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining THAP-family protein and/or nucleic acid expression as well as THAP-family activity, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant THAP-family expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with a THAP-family protein, nucleic acid expression or activity. For example, mutations in a THAP-family gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with a THAP-family protein, nucleic acid expression or activity.

Accordingly, the methods of the present invention are applicable generally to diseases related to regulation of apoptosis, including but not limited to disorders characterized by unwanted cell proliferation or generally aberrant control of differentiation, for example neoplastic or hyperplastic disorders, as well as disorders related to proliferation or lack thereof of endothelial cells, inflammatory disorders and neurodegenerative disorders.

#### *Diagnostic Assays*

An exemplary method for detecting the presence (quantitative or not) or absence of a THAP-family protein or nucleic acid in a biological sample involves obtaining a biological sample

from a test subject and contacting the biological sample with a compound or an agent capable of detecting a THAP-family protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes THAP-family protein such that the presence of the THAP-family protein or nucleic acid is detected in the biological sample. A preferred agent for detecting a THAP-family mRNA or genomic DNA is a  
5 labeled nucleic acid probe capable of hybridizing to a THAP-family mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length THAP-family nucleic acid, such as the nucleic acid of SEQ ID NO: 160 such as a nucleic acid of at least 15, 30, 50, 100, 250, 400, 500 or 1000 nucleotides in length and sufficient to specifically hybridize under stringent conditions to a THAP-family mRNA or genomic DNA or a portion of a THAP-family nucleic acid. Other suitable probes  
10 for use in the diagnostic assays of the invention are described herein.

In preferred embodiments, the subject method can be characterized by generally comprising detecting, in a tissue sample of the subject (e.g. a human patient), the presence or absence of a genetic lesion characterized by at least one of (i) a mutation of a gene encoding one of the subject THAP-family proteins or (ii) the mis-expression of a THAP-family gene. To illustrate, such genetic  
15 lesions can be detected by ascertaining the existence of at least one of (i) a deletion of one or more nucleotides from a THAP-family gene, (ii) an addition of one or more nucleotides to such a THAP-family gene, (iii) a substitution of one or more nucleotides of a THAP-family gene, (iv) a gross chromosomal rearrangement or amplification of a THAP-family gene, (v) a gross alteration in the level of a messenger RNA transcript of a THAP-family gene, (vi) aberrant modification of a THAP-  
20 family gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a THAP-family gene, and (viii) a non-wild type level of a THAP-family -target protein.

A preferred agent for detecting a THAP-family protein is an antibody capable of binding to a THAP-family protein, preferably an antibody with a detectable label. Antibodies can be  
25 polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')<sub>2</sub>) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include  
30 detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect a THAP-family mRNA, protein, or genomic DNA in a  
35 biological sample in vitro as well as in vivo. For example, in vitro techniques for detection of a THAP-family mRNA include Northern hybridizations and in situ hybridizations. In vitro techniques for detection of a THAP-family protein include enzyme linked immunosorbent assays (ELISAs),



Western blots, immunoprecipitations and immunofluorescence. In vitro techniques for detection of a THAP-family genomic DNA include Southern hybridizations. Furthermore, in vivo techniques for detection of a THAP-family protein include introducing into a subject a labeled anti-THAP-family antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In yet another exemplary embodiment, aberrant methylation patterns of a THAP-family gene can be detected by digesting genomic DNA from a patient sample with one or more restriction endonucleases that are sensitive to methylation and for which recognition sites exist in the THAP-family gene (including in the flanking and intronic sequences). See, for example, Buiting et al. (1994) Human Mol Genet 3:893-895. Digested DNA is separated by gel electrophoresis, and hybridized with probes derived from, for example, genomic or cDNA sequences. The methylation status of the THAP-family gene can be determined by comparison of the restriction pattern generated from the sample DNA with that for a standard of known methylation.

Furthermore, gene constructs such as those described herein can be utilized in diagnostic assays to determine if a cell's growth or differentiation state is no longer dependent on the regulatory function of a THAP-family protein, e.g. in determining the phenotype of a transformed cell. Such knowledge can have both prognostic and therapeutic benefits. To illustrate, a sample of cells from the tissue can be obtained from a patient and dispersed in appropriate cell culture media, a portion of the cells in the sample can be caused to express a recombinant THAP-family protein or a THAP-family target protein, e.g. by transfection with a expression vector described herein, or to increase the expression or activity of an endogenous THAP-family protein or THAP-family target protein, and subsequent growth of the cells assessed. The absence of a change in phenotype of the cells despite expression of the THAP-family or THAP-family target protein may be indicative of a lack of dependence on cell regulatory pathways which includes the THAP-family or THAP-family target protein, e.g. THAP-family- or THAP-family target-mediated transcription. Depending on the nature of the tissue of interest, the sample can be in the form of cells isolated from, for example, a blood sample, an exfoliated cell sample, a fine needle aspirant sample, or a biopsied tissue sample. Where the initial sample is a solid mass, the tissue sample can be minced or otherwise dispersed so that cells can be cultured, as is known in the art.

In yet another embodiment, a diagnostic assay is provided which detects the ability of a THAP-family gene product, e.g., isolated from a biopsied cell, to bind to other cellular proteins. For instance, it will be desirable to detect THAP-family mutants which, while expressed at appreciable levels in the cell, are defective at binding a THAP-family target protein (having either diminished or enhanced binding affinity). Such mutants may arise, for example, from mutations, e.g., point mutants, which may be impractical to detect by the diagnostic DNA sequencing techniques or by the immunoassays described above. The present invention accordingly further contemplates diagnostic screening assays which generally comprise cloning one or more THAP-family genes

from the sample cells, and expressing the cloned genes under conditions which permit detection of an interaction between that recombinant gene product and a target protein, e.g., for example the THAP1 gene and a target PAR4 protein or a PML-NB protein. As will be apparent from the description of the various drug screening assays set forth below, a wide variety of techniques can be used to determine the ability of a THAP-family protein to bind to other cellular components. These techniques can be used to detect mutations in a THAP-family gene which give rise to mutant proteins with a higher or lower binding affinity for a THAP-family target protein relative to the wild-type THAP-family. Conversely, by switching which of the THAP-family target protein and THAP-family protein is the "bait" and which is derived from the patient sample, the subject assay can also be used to detect THAP-family target protein mutants which have a higher or lower binding affinity for a THAP-family protein relative to a wild type form of that THAP-family target protein.

In an exemplary embodiment, a PAR4 or a PMB-NB protein (e.g. wild-type) can be provided as an immobilized protein (a "target"), such as by use of GST fusion proteins and glutathione treated microtitre plates. A THAP1 gene (a "sample" gene) is amplified from cells of a patient sample, e.g., by PCR, ligated into an expression vector, and transformed into an appropriate host cell. The recombinantly produced THAP1 protein is then contacted with the immobilized PAR4 or PMB-NB protein, e.g., as a lysate or a semi-purified preparation, the complex washed, and the amount of PAR4 or PMB-NB protein /THAP1 complex determined and compared to a level of wild-type complex formed in a control. Detection can be by, for instance, an immunoassay using antibodies against the wild-type form of the THAP1 protein, or by virtue of a label provided by cloning the sample THAP1 gene into a vector which provides the protein as a fusion protein including a detectable tag. For example, a myc epitope can be provided as part of a fusion protein with the sample THAP1 gene. Such fusion proteins can, in addition to providing a detectable label, also permit purification of the sample THAP1 protein from the lysate prior to application to the immobilized target. In yet another embodiment of the subject screening assay, the two hybrid assay, described in the appended examples, can be used to detect mutations in either a THAP-family gene or THAP-family target gene which alter complex formation between those two proteins.

Accordingly, the present invention provides a convenient method for detecting mutants of THAP-family genes encoding proteins which are unable to physically interact with a THAP-family target "bait" protein, which method relies on detecting the reconstitution of a transcriptional activator in a THAP-family/THAP-family target-dependent fashion.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a serum sample isolated by conventional means from a subject. In another embodiment, the methods further involve obtaining

a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting a THAP-family protein, mRNA, or genomic DNA, such that the presence of a THAP-family protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of a THAP-family protein, mRNA or genomic DNA in the control sample with the presence of a THAP-family protein, mRNA or genomic DNA in the test sample. The invention also encompasses kits for detecting the presence of THAP-family protein, mRNA or genomic DNA in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting a THAP-family protein or mRNA or genomic DNA in a biological sample; means for determining the amount of a THAP-family member in the sample; and means for comparing the amount of THAP-family member in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect THAP-family protein or nucleic acid.

In certain embodiments, detection involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegren et al. (1988) *Science* 241:1077-1080; and Nakazawa et al. (1994) *PNAS* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the THAP-family-gene (see Abravaya et al. (1995) *Nucleic Acids Res.* 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a THAP-family gene under conditions such that hybridization and amplification of the THAP-family-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Genotyping assays for diagnostics generally require the previous amplification of the DNA region carrying the biallelic marker of interest. However, ultrasensitive detection methods which do not require amplification are also available. Methods well-known to those skilled in the art that can be used to detect biallelic polymorphisms include methods such as, conventional dot blot analyzes, single strand conformational polymorphism analysis (SSCP) described by Orita et al., *PNAS* 86: 2766-2770 (1989), denaturing gradient gel electrophoresis (DGGE), heteroduplex analysis, mismatch cleavage detection, and other conventional techniques as described in Sheffield et al. (1991), White et al. (1992), and Grompe et al. (1989 and 1993) (Sheffield, V.C. et al, *Proc. Natl. Acad. Sci. U.S.A* 49:699-706 (1991); White, M.B. et al., *Genomics* 12:301-306 (1992); Grompe, M. et al., *Proc. Natl. Acad. Sci. U.S.A* 86:5855-5892 (1989); and Grompe, M. *Nature Genetics* 5:111-117 (1993)). Another method for determining the identity of the nucleotide present

at a particular polymorphic site employs a specialized exonuclease-resistant nucleotide derivative as described in U.S. patent 4,656,127. Further methods are described as follows.

The nucleotide present at a polymorphic site can be determined by sequencing methods. In a preferred embodiment, DNA samples are subjected to PCR amplification before sequencing as described above. DNA sequencing methods are described in "Sequencing Of Amplified Genomic DNA And Identification Of Single Nucleotide Polymorphisms". Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. Sequence analysis allows the identification of the base present at the biallelic marker site.

In microsequencing methods, the nucleotide at a polymorphic site in a target DNA is detected by a single nucleotide primer extension reaction. This method involves appropriate microsequencing primers which, hybridize just upstream of the polymorphic base of interest in the target nucleic acid. A polymerase is used to specifically extend the 3' end of the primer with one single ddNTP (chain terminator) complementary to the nucleotide at the polymorphic site. Next the identity of the incorporated nucleotide is determined in any suitable way. Typically, microsequencing reactions are carried out using fluorescent ddNTPs and the extended microsequencing primers are analyzed by electrophoresis on ABI 377 sequencing machines to determine the identity of the incorporated nucleotide as described in EP 412 883. Alternatively capillary electrophoresis can be used in order to process a higher number of assays simultaneously. Different approaches can be used for the labeling and detection of ddNTPs. A homogeneous phase detection method based on fluorescence resonance energy transfer has been described by Chen and Kwok (1997) and, Chen and Kwok (*Nucleic Acids Research* 25:347-353 1997) and Chen et al. (*Proc. Natl. Acad. Sci. USA* 94/20 10756-10761,1997)). In this method, amplified genomic DNA fragments containing polymorphic sites are incubated with a 5'-fluorescein-labeled primer in the presence of allelic dye-labeled dideoxyribonucleoside triphosphates and a modified Taq polymerase. The dye-labeled primer is extended one base by the dye-terminator specific for the allele present on the template. At the end of the genotyping reaction, the fluorescence intensities of the two dyes in the reaction mixture are analyzed directly without separation or purification. All these steps can be performed in the same tube and the fluorescence changes can be monitored in real time. Alternatively, the extended primer may be analyzed by MALDI-TOF Mass Spectrometry. The base at the polymorphic site is identified by the mass added onto the microsequencing primer (see Haff and Smirnov, 1997, *Genome Research*, 7:378-388, 1997). In another example, Pastinen et al., (*Genome Research* 7:606-614, 1997)) describe a method for multiplex detection of single nucleotide polymorphism in which the solid phase minisequencing principle is applied to an oligonucleotide array format. High-density arrays of DNA probes attached to a solid support (DNA chips) are further described below.

Other assays include mismatch detection assays, based on the specificity of polymerases and ligases. Polymerization reactions places particularly stringent requirements on correct base pairing of the 3' end of the amplification primer and the joining of two oligonucleotides hybridized to a target DNA sequence is quite sensitive to mismatches close to the ligation site, especially at the 3' end.

A preferred method of determining the identity of the nucleotide present at an allele involves nucleic acid hybridization. Any hybridization assay may be used including Southern hybridization, Northern hybridization, dot blot hybridization and solid-phase hybridization (see Sambrook et al., *Molecular Cloning - A Laboratory Manual*, Second Edition, Cold Spring Harbor Press, N.Y., 1989)). Hybridization refers to the formation of a duplex structure by two single stranded nucleic acids due to complementary base pairing. Hybridization can occur between exactly complementary nucleic acid strands or between nucleic acid strands that contain minor regions of mismatch. Specific probes can be designed that hybridize to one form of a biallelic marker and not to the other and therefore are able to discriminate between different allelic forms. Allele-specific probes are often used in pairs, one member of a pair showing perfect match to a target sequence containing the original allele and the other showing a perfect match to the target sequence containing the alternative allele. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Stringent, sequence specific hybridization conditions, under which a probe will hybridize only to the exactly complementary target sequence are well known in the art (Sambrook et al., 1989). The detection of hybrid duplexes can be carried out by a number of methods. Various detection assay formats are well known which utilize detectable labels bound to either the target or the probe to enable detection of the hybrid duplexes. Typically, hybridization duplexes are separated from unhybridized nucleic acids and the labels bound to the duplexes are then detected. Further, standard heterogeneous assay formats are suitable for detecting the hybrids using the labels present on the primers and probes. (see Landegren U. et al., *Genome Research*, 8:769-776, 1998).

Hybridization assays based on oligonucleotide arrays rely on the differences in hybridization stability of short oligonucleotides to perfectly matched and mismatched target sequence variants. Efficient access to polymorphism information is obtained through a basic structure comprising high-density arrays of oligonucleotide probes attached to a solid support (e.g., the chip) at selected positions. Chips of various formats for use in detecting biallelic polymorphisms can be produced on a customized basis by Affymetrix (GeneChip), Hyseq (HyChip and HyGnostics), and Protogene Laboratories.

In general, these methods employ arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from an individual which, target sequences include a polymorphic marker. EP 785280, describes a tiling strategy for the detection of single nucleotide

polymorphisms. Briefly, arrays may generally be "tiled" for a large number of specific polymorphisms, further described in PCT application No. WO 95/11995. Upon completion of hybridization with the target sequence and washing of the array, the array is scanned to determine the position on the array to which the target sequence hybridizes. The hybridization data from the scanned array is then analyzed to identify which allele or alleles of the biallelic marker are present in the sample. Hybridization and scanning may be carried out as described in PCT application No. WO 92/10092 and WO 95/11995 and US patent No. 5,424,186. Solid supports and polynucleotides of the present invention attached to solid supports are further described in "Oligonucleotide Probes And Primers".

## 10 DETECTING CHEMOKINES

Some aspects of the present invention relate to the detection of chemokines by contacting a chemokine or a sample containing a chemokine with a THAP-type chemokine-binding agent. In some embodiments, the chemokines or the THAP-type chemokine-binding agents are labeled. Many labels and methods of conjugating such labels to a chemokine or a THAP-type chemokine-binding agent are known in the art. Additionally, labeled molecules, such as antibodies, which have an affinity for a THAP-type chemokine-binding agent can be used to detect the chemokine that is bound to a THAP-type chemokine-binding agent using a number of assay formats that are well known in the art.

An exemplary method for detecting the presence (quantitative or not) or absence of a chemokine, including, but not limited to, a chemokine in a biological sample, involves obtaining a chemokine or a sample containing a chemokine and contacting it with a compound or an agent capable of detecting the chemokine. In some embodiments, such an agent is a THAP-type chemokine-binding agent. Chemokines which can be detected using a method that employs a THAP-type chemokine-binding agent include, but are not limited to, XCL1, XCL2, CCL1, CCL2, CCL3, CCL3L1, SCYA3L2, CCL4, CCL4L, CCL5, CCL6, CCL7, CCL8, SCYA9, SCYA10, CCL11, SCYA12, CCL13, CCL14, CCL15, CCL16, CCL17, CCL18, CCL19, CCL20, CCL21, CCL22, CCL23, CCL24, CCL25, CCL26, CCL27, CCL28, clone 391, CARP CC-1, CCL1, CK-1, regakine-1, K203, CXCL1, CXCL1P, CXCL2, CXCL3, PF4, PF4V1, CXCL5, CXCL6, PPBP, SPBPBP, IL8, CXCL9, CXCL10, CXCL11, CXCL12, CXCL14, CXCL15, CXCL16, NAP-4, LFCA-1, Scyba, JSC, VHSV-induced protein, CX3CL1 and fCL1.

In some embodiments, the detection method comprises detecting, in a biological sample, such as a tissue or fluid sample from a subject (such as, a human patient), the presence or absence of a chemokine by contacting the biological sample with a THAP-type chemokine-binding agent and detecting a complex between the chemokine and the THAP-type chemokine-binding agent or detecting a THAP-type chemokine-binding agent which was previously bound to the chemokine but which has been released from the chemokine.

In some embodiments of the present invention, the THAP-type chemokine-binding agent is labeled directly. In other embodiments, the THAP-type chemokine-binding agent is detected using a labeled antibody having affinity for the THAP-type chemokine-binding agent. Such antibodies may directly carry the detectable label or be recognized by a labeled second antibody. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')<sub>2</sub>) can be used. The term "labeled", with regard to the antibody or other detectable molecule, is intended to encompass direct labeling of the antibody or molecule by coupling (i.e., physically linking) a detectable substance to the antibody or molecule, as well as indirect labeling of the antibody or molecule by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a THAP-type chemokine-binding agent with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. Accordingly, the detection method can be used to detect a chemokine in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of a chemokine include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vivo* techniques for detection of a chemokine include introducing into a subject a labeled THAP-type chemokine-binding agent. For example, the THAP-type chemokine-binding agent can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

Other aspects of the present invention relate to a system for chemokine detection. Such a chemokine detection system comprises a THAP-type chemokine-binding agent bound to a solid support. A number of adequate solid support materials are known in the art and include, but are not limited to, cellulose, nylon or other polymer backings, plastics such as microtiter plates, synthetic beads and resins such as sepharose, glass, magnetic beads, latex particles, sheep (or other animal) red blood cells, duracytes and others. Suitable methods for immobilizing the THAP-type chemokine-binding agent to the solid support are well known in the art.

Some embodiments of the present invention relate to kits which comprise a THAP-type chemokine-binding agent and instructions which describe detecting or inhibiting chemokines with the THAP-type chemokine-binding agent. For example, the kit includes an ampule of THAP-type chemokine-binding agent that is stored so as to prevent damage or inactivation of the agent upon prolonged storage. Such methods can include, but are not limited to, lyophilization and freezing in an appropriate buffer. The kit also can contain chemokines to serve as a positive control sample when the kit is used for chemokine binding, detection or inhibition.

In some embodiments of the present invention, kits are packaged containing a heterogeneous mixture of THAP-type chemokine-binding agents, wherein each of the agents has a different affinity for one or more chemokines. Alternatively, some kits comprise a panel of THAP-

type chemokine-binding agents, wherein each THAP-type chemokine binding agent has a different affinity for a particular chemokine. For example, the kit can comprise a panel of three THAP-type chemokine-binding agents, wherein the first agent has a high affinity for SLC but a low affinity for CXCL9, the second agent has a moderate affinity for both SLC and CXCL9, and the third agent has a low affinity for SLC and a high affinity for CXCL9. Panels of THAP-type chemokine-binding agents can be larger or small than that exemplified above and the number and types of chemokines that are detected can be more or less than that exemplified above. Kits containing such panels of THAP-type chemokine-binding agents can be used to reliably distinguish mixed samples of chemokines. Additionally, such panels can be used to bind or inhibit multiple different chemokines in a mixed chemokine sample.

Having generally described this invention, a further understanding can be obtained by reference to certain specific examples which are provided herein for purposes of illustration only, and are not intended to be limiting unless otherwise specified.

## EXAMPLES

### EXAMPLE 1

#### Isolation of the THAP1 cDNA in a two-hybrid screen with chemokine SLC/CCL21

In an effort to define the function of novel HEVEC proteins and the cellular pathways involved, we used different baits to screen a two-hybrid cDNA library generated from microvascular human HEV endothelial cells (HEVEC). HEVEC were purified from human tonsils by immunomagnetic selection with monoclonal antibody MECA-79 as previously described (Girard and Springer (1995) Immunity 2:113-123). The SMART PCR cDNA library Construction Kit (Clontech, Palo Alto, CA, USA) was first used to generate full-length cDNAs from 1 µg HEVEC total RNA. Oligo-dT-primed HEVEC cDNA were then digested with SfiI and directionally cloned into pGAD424-Sfi, a two-hybrid vector generated by inserting a SfiI linker (5'-GAATTCGGCCATTATGGCCTGCAGGATCCGGCCGCCTCGGCCAGGATCC-3') (SEQ ID NO: 181) between EcoRI and BamHI cloning sites of pGAD424 (Clontech). The resulting pGAD424-HEVEC cDNA two-hybrid library (mean insert size > 1 kb, ~ 3x10<sup>6</sup> independent clones) was amplified in *E. coli*. To identify potential protein partners of chemokine SLC/6CKine, screening of the two-hybrid HEVEC cDNA library was performed using as bait a cDNA encoding the mature form of human SLC/CCL21 (amino acids 24-134, GenBank Accession No: NP\_002980, SEQ ID NO: 182), amplified by PCR from HEVEC RNA with primers hSLC.5' (5'-GCGGGATCCGTAGTGATGGAGGGGCTCAGGACTGTTG-3') (SEQ ID NO: 183) and hSLC.3' (5'-GCGGGATCCCTATGGCCCTTTAGGGGTCTGTGACC-3') (SEQ ID NO: 184), digested with BamHI and inserted into the BamHI cloning site of MATCHMAKER two-hybrid system 2 vector pGBT9 (Clontech). Briefly, pGBT9-SLC was cotransformed with the pGAD424-HEVEC cDNA library in yeast strain Y190 (Clontech). 1.5x10<sup>7</sup> yeast transformants were screened



and positive protein interactions were selected by His auxotrophy. The plates were incubated at 30°C for 5 days. Plasmid DNA was extracted from positive colonies and used to verify the specificity of the interaction by cotransformation in AH109 with pGBT9-SLC or control baits pGBT9, pGBT9-lamin. Eight independent clones isolated in this two-hybrid screen were characterized. They were found to correspond to a unique human cDNA encoding a novel human protein of 213 amino acids, designated THAP1, that exhibits 93% identity with its mouse orthologue (Figure 1A). The only noticeable motifs in the THAP1 predicted protein sequence were a short proline-rich domain in the middle part and a consensus nuclear localization sequence (NLS) in the carboxy terminal part (Figure 1B). Databases searches with the THAP1 sequence failed to reveal any significant similarity to previously characterized proteins with the exception of the first 90 amino acids that may define a novel protein motif associated with apoptosis, hereafter referred to as THAP domain (see Figure 1B, Figures 9A-9C, and Figure 10).

## EXAMPLE 2

### Northern Blot

To determine the tissue distribution of THAP1 mRNA, we performed Northern blot analysis of 12 different adult human tissues (Fig 2). Multiple Human Tissues Northern Blots (CLONTECH) were hybridized according to manufacturer's instructions. The probe was a PCR product corresponding to the THAP1 ORF, <sup>32</sup>P-labeled with the Prime-a-Gene Labeling System (PROMEGA). A 2.2-kb mRNA band was detected in brain, heart, skeletal muscle, kidney, liver, and placenta. In addition to the major 2.2 kb band, lower molecular weight bands were detected, that are likely to correspond to alternative splicing or polyadenylation of the THAP1 pre-mRNA. The presence of THAP1 mRNAs in many different tissues suggests that THAP1 has a widespread, although not ubiquitous, tissue distribution in the human body.

## EXAMPLE 3

### Analysis of the subcellular THAP1 localization

To analyze the subcellular localization of the THAP1 protein, the THAP1 cDNA was fused to the coding sequence of GFP (Green Fluorescent Protein). The full-length coding region of THAP1 was amplified by PCR from HEVEC cDNA with primers 2HMR10 (5'-CCGAATTCAGGATGGTGCAGTCCTGCTCCGCCT-3') (SEQ ID NO: 185) and 2HMR9 (5'-CGCGGATCCTGCTGGTACTTCAACTATTTCAAAGTAGTC-3') (SEQ ID NO: 186), digested with EcoRI and BamHI, and cloned in frame downstream of the Enhanced Green Fluorescent Protein (EGFP) ORF in pEGFP.C2 vector (Clontech) to generate pEGFP.C2-THAP1. The GFP/THAP1 expression construct was then transfected into human primary endothelial cells from umbilical vein (HUVEC, PromoCell, Heidelberg, Germany). HUVEC were grown in complete ECGM medium (PromoCell, Heidelberg, Germany), plated on coverslips and transiently transfected in RPMI medium using GeneJammer transfection reagent according to manufacturer instructions (Stratagene, La Jolla, CA, USA). Analysis by fluorescence microscopy 24h later

revealed that the GFP/THAP1 fusion protein localizes exclusively in the nucleus with both a diffuse distribution and an accumulation into speckles while GFP alone exhibits only a diffuse staining over the entire cell. To investigate the identity of the speckled domains with which GFP/THAP1 associates, we used indirect immunofluorescence microscopy to examine a possible colocalization of the nuclear dots containing GFP/THAP1 with known nuclear domains (replication factories, splicing centers, nuclear bodies).

Cells transfected with GFP-tagged expression constructs were allowed to grow for 24 h to 48 h on coverslips. Cells were washed twice with PBS, fixed for 15 min at room temperature in PBS containing 3.7% formaldehyde, and washed again with PBS prior to neutralization with 50mM NH<sub>4</sub>Cl in PBS for 5 min at room temperature. Following one more PBS wash, cells were permeabilized 5 min at room temperature in PBS containing 0.1% Triton-X100, and washed again with PBS. Permeabilized cells were then blocked with PBS-BSA (PBS with 1% bovine serum albumin) for 10 min and then incubated 2 hr at room temperature with the following primary antibodies diluted in PBS-BSA: rabbit polyclonal antibodies against human Daxx (1/50, M-112, Santa Cruz Biotechnology) or mouse monoclonal antibodies anti-PML (mouse IgG1, 1/30, mAb PG-M3 from Dako, Glostrup, Denmark). Cells were then washed three times 5 min at room temperature in PBS-BSA, and incubated for 1 hr with Cy3 (red fluorescence)-conjugated goat anti-mouse or anti-rabbit IgG (1/1000, Amersham Pharmacia Biotech) secondary antibodies, diluted in PBS-BSA. After extensive washing in PBS, samples were air dried and mounted in Mowiol. Images were collected on a Leica confocal laser scanning microscope. The GFP (green) and Cy3 (red) fluorescence signals were recorded sequentially for identical image fields to avoid cross-talk between the channels.

This analysis revealed that GFP-THAP1 staining exhibits a complete overlap with the staining pattern obtained with antibodies directed against PML. The colocalization of GFP/THAP1 and PML was observed both in nuclei with few PML-NBs (less than ten) and in nuclei with a large number of PML-NBs. Indirect immunofluorescence staining with antibodies directed against Daxx, another well characterized component of PML-NBs, was performed to confirm the association of GFP/THAP1 with PML-NBs. We found a complete colocalization of GFP/THAP1 and Daxx in PML-NBs. Together, these results reveal that THAP1 is a novel protein associated with PML-NBs.

#### EXAMPLE 4

##### Identification of proteins interacting with THAP1 in human HEVECs: two-hybrid assay

##### **THAP1 forms a complex with the pro-apoptotic protein PAR4**

To identify potential protein partners of THAP1, screening of the two-hybrid HEVEC cDNA library was performed using as a bait the human THAP1 full length cDNA inserted into the MATCHMAKER two-hybrid system 3 vector pGBKT7 (Clontech). Briefly, the full-length coding region of THAP1 was amplified by PCR from HEVEC cDNA with primers 2HMR10 (5'-CCGAATTCAGGATGGTGCAGTCCTGCTCCGCCT-3') (SEQ ID NO: 187) and 2HMR9 (5'-

CGCGGATCCTGCTGGTACTTCAACTATTTCAAAGTAGTC-3') (SEQ ID NO: 188), digested with EcoRI and BamHI, and cloned in frame downstream of the Gal4 Binding Domain (Gal4-BD) in pGBKT7 vector to generate pGBKT7-THAP1. pGBKT7-THAP1 was then cotransformed with the pGAD424-HEVEC cDNA library in yeast strain AH109 (Clontech). 1.5x10<sup>7</sup> yeast transformants were screened and positive protein interactions were selected by His and Ade double auxotrophy according to manufacturer's instructions (MATCHMAKER two-hybrid system 3, Clontech). The plates were incubated at 30°C for 5 days. Plasmid DNA was extracted from these positive colonies and used to verify the specificity of the interaction by cotransformation in AH109 with pGBKT7-THAP1 or control baits pGBKT7, pGBKT7-lamin and pGBKT7-hevin. Three clones which specifically interacted with THAP1 were obtained in the screen; sequencing of these clones revealed three identical library plasmids that corresponded to a partial cDNA coding for the last 147 amino acids (positions 193-342) of the human pro-apoptotic protein PAR4 (Fig 3A). Positive interaction between THAP1 and Par4 was confirmed using full length Par4 bait (pGBKT-Par4) and prey (pGADT7-Par4). Full-length human Par4 was amplified by PCR from human thymus cDNA (Clontech), with primers *Par4.8* (5'-GCGGAATTCATGGCGACCGGTGGCTACCGGACC-3') (SEQ ID NO: 189) and *Par4.5* (5'-GCGGGATCCCTCTACCTGGTCAGCTGACCCACAAC-3') (SEQ ID NO: 190), digested with EcoRI and BamHI, and cloned in pGBKT7 and pGADT7 vectors, to generate pGBKT7-Par4 and pGADT7-Par4. Positive interaction between THAP1 and Par4 was confirmed by cotransformation of AH109 with pGBKT7-THAP1 and pGADT7-Par4 or pGBKT7-Par4 and pGADT7-THAP1 and selection of transformants by His and Ade double auxotrophy according to manufacturer's instructions (MATCHMAKER two-hybrid system 3, Clontech). To generate pGADT7-THAP1, the full-length coding region of THAP1 was amplified by PCR from HEVEC cDNA with primers *2HMR10* (5'-CCGAATTCAGGATGGTGCAGTCCTGCTCCGCCT-3') (SEQ ID NO: 191) and *2HMR9* (5'-CGCGGATCCTGCTGGTACTTCAACTATTTCAAAGTAGTC-3') (SEQ ID NO: 192), digested with EcoRI and BamHI, and cloned in frame downstream of the Gal-4 Activation Domain (Gal4-AD) in pGADT7 two-hybrid vector (Clontech).

We then examined whether the leucine zipper/death domain at the C-terminus of Par4, previously shown to be involved in Par4 binding to WT-1 and aPKC, was required for the interaction between THAP1 and Par4. Two Par4 mutants were constructed for that purpose, Par4Δ and Par4DD. Par4Δ lacks the leucine zipper/death domain while Par4DD contains this domain. pGBKT7-Par4Δ(amino acids 1-276) and pGADT7-Par4Δ were constructed by sub-cloning a EcoRI-BglII fragment from pGADT7-Par4 into the EcoRI and BamHI sites of pGBKT7 and pGADT7. Par4DD (amino acids 250-342) was amplified by PCR, using pGBKT7-Par4 as template, with primers *Par4.4* (5'-CGCGAATTCGCCATCATGGGGTTCCTAGATATAACAGGGATGCAA-3') (SEQ ID NO: 193) and *Par4.5*, and cloned into the EcoRI and BamHI sites of pGBKT7 and pGADT7 to obtain

pGBKT7-Par4DD and pGADT7-Par4DD. Two-hybrid interaction between THAP1 and Par4 mutants was tested by cotransformation of AH109 with pGBKT7-THAP1 and pGADT7-Par4 $\Delta$  or pGADT7-Par4DD and selection of transformants by His and Ade double auxotrophy according to manufacturer's instructions (MATCHMAKER two-hybrid system 3, Clontech). We found that the Par4 leucine zipper/death domain (Par4DD) is not only required but also sufficient for the interaction with THAP1 (Fig 3A). Similar results were obtained when two-hybrid experiments were performed in the opposite orientation using Par4 or Par4 mutants (Par4 $\Delta$  and Par4DD) as baits instead of THAP1 (Fig 3A).

## EXAMPLE 5

In vitro THAP1/Par4 interaction assay

To confirm the interaction observed in yeast, we performed *in vitro* GST pull down assays. Par4DD, expressed as a GST-tagged fusion protein and immobilized on glutathione sepharose, was incubated with radiolabeled *in vitro* translated THAP1. To generate the GST-Par4DD expression vector, Par4DD (amino acids 250-342) was amplified by PCR with primers *Par4.10* (5'-GCCGGATCCGGGTTCCTAGATATAACAGGGATGCAA-3') (SEQ ID NO: 194) and *Par4.5*, and cloned in frame downstream of the Glutathion S-Transferase ORF, into the BamHI site of the pGEX-2T prokaryotic expression vector (Amersham Pharmacia Biotech, Saclay, France). GST-Par4DD(amino acids 250-342) fusion protein encoded by plasmid pGEX-2T-Par4DD and control GST protein encoded by plasmid pGEX-2T, were then expressed in *E.Coli* DH5 $\alpha$  and purified by affinity chromatography with glutathione sepharose according to supplier's instructions (Amersham Pharmacia Biotech). The yield of proteins used in GST pull-down assays was determined by SDS-Polyarylamide Gel Electrophoresis (PAGE) and Coomassie blue staining analysis. *In vitro* translated THAP1 was generated with the TNT-coupled reticulocyte lysate system (Promega, Madison, WI, USA) using pGBKT7-THAP1 vector as template. 25  $\mu$ l of  $^{35}$ S-labelled wild-type THAP1 was incubated with immobilized GST-Par4 or GST proteins overnight at 4  $^{\circ}$ C, in the following binding buffer: 10 mM NaPO $_4$  pH 8.0, 140 mM NaCl, 3 mM MgCl $_2$ , 1mM dithiothreitol (DTT), 0.05% NP40, and 0.2 mM phenylmethyl sulphonyl fluoride (PMSF), 1 mM Na Vanadate, 50mM  $\beta$  Glycerophosphate, 25  $\mu$ g/ml chymotrypsine, 5  $\mu$ g/ml aprotinin, 10  $\mu$ g/ml Leupeptin. Beads were then washed 5 times in 1 ml binding buffer. Bound proteins were eluted with 2X Laemmli SDS-PAGE sample buffer, fractionated by 10% SDS-PAGE and visualized by fluorography using Amplify (Amersham Pharmacia Biotech). As expected, GST/Par4DD interacted with THAP1 (Fig 3B). In contrast, THAP1 failed to interact with GST beads.

## EXAMPLE 6

In vivo THAP1/Par4 interaction assay

To provide further evidence for a physiological interaction between THAP1 and Par4 *in vivo* interactions between THAP1 and PAR4 were investigated. For that purpose, confocal immunofluorescence microscopy was used to analyze the subcellular localization of epitope-tagged

Par4DD in primary human endothelial cells transiently cotransfected with pEF-*myc*Par4DD eukaryotic expression vector and GFP or GFP-THAP1 expression vectors (pEGFP.C2 and pEGFP.C2-THAP1, respectively). To generate pEF-*myc*Par4DD, *myc*Par4DD (amino acids 250-342) was amplified by PCR using pGBT7-Par4DD as template, with primers *myc*.BD7 (5'-GCGCTCTAGAGCCATCATGGAGGAGCAGAAGCTGATC-3') (SEQ ID NO: 195) and *Par4.9* (5'-CTTGCGGCCGCTCTACCTGGTCAGCTGACCCACAAC-3') (SEQ ID NO: 196), and cloned into the XbaI and NotI sites of the pEF-BOS expression vector (Mizushima and Nagata, Nucleic Acids Research, 18:5322, 1990). Primary human endothelial cells from umbilical vein (HUVEC, PromoCell, Heidelberg, Germany) were grown in complete ECGM medium (PromoCell, Heidelberg, Germany), plated on coverslips and transiently transfected in RPMI medium using GeneJammer transfection reagent according to manufacturer instructions (Stratagene, La Jolla, CA, USA). Cells co-transfected with pEF-*myc*Par4DD and GFP-tagged expression constructs were allowed to grow for 24 h to 48 h on coverslips. Cells were washed twice with PBS, fixed for 15 min at room temperature in PBS containing 3.7% formaldehyde, and washed again with PBS prior to neutralization with 50mM NH<sub>4</sub>Cl in PBS for 5 min at room temperature. Following one more PBS wash, cells were permeabilized 5 min at room temperature in PBS containing 0.1% Triton-X100, and washed again with PBS. Permeabilized cells were then blocked with PBS-BSA (PBS with 1% bovine serum albumin) for 10 min and then incubated 2 hr at room temperature with mouse monoclonal antibody anti-*myc* epitope (mouse IgG1, 1/200, Clontech) diluted in PBS-BSA. Cells were then washed three times 5 min at room temperature in PBS-BSA, and incubated for 1 hr with Cy3 (red fluorescence)-conjugated goat anti-mouse (1/1000, Amersham Pharmacia Biotech) secondary antibodies, diluted in PBS-BSA. After extensive washing in PBS, samples were air dried and mounted in Mowiol. Images were collected on a Leica confocal laser scanning microscope. The GFP (green) and Cy3 (red) fluorescence signals were recorded sequentially for identical image fields to avoid cross-talk between the channels.

In cells transiently co-transfected with pEF-*myc*Par4DD and GFP expression vector, ectopically expressed *myc*-Par4DD was found to accumulate both in the cytoplasm and the nucleus of the majority of the cells. In contrast, transient cotransfection of pEF-*myc*Par4DD and GFP-THAP1 expression vectors dramatically shifted *myc*-Par4DD from a diffuse cytosolic and nuclear localization to a preferential association with PML-NBs. The effect of GFP-THAP1 on *myc*-Par4DD localization was specific since it was not observed with GFP-APS kinase-1 (APSK-1), a nuclear enzyme unrelated to THAP1 and apoptosis [Besset et al., FASEB J, 14:345-354, 2000]. This later result shows that GFP-THAP1 recruits *myc*-Par4DD at PML-NBs and provides *in vivo* evidence for a direct interaction of THAP1 with the pro-apoptotic protein Par4.

35

## EXAMPLE 7

Identification of a novel arginine-rich Par4 binding motif

To identify the sequences mediating THAP1 binding to Par4, a series of THAP1 deletion constructs was generated. Both amino-terminal (THAP1-C1, -C2, -C3) and carboxy-terminal (THAP1-N1, -N2, -N3) deletion mutants (Figure 4A) were amplified by PCR using plasmid pEGFP.C2-THAP1 as a template and the following primers: *2HMR12* (5'-GCGGAATTCAAAGAAGATCTTCTGGAGCCACAGGAAC-3') (SEQ ID NO: 197) and *2HMR9* (5'-CGCGGATCCTGCTGGTACTTCAACTATTTCAAAGTAGTC-3') (SEQ ID NO: 198) for THAP1-C1 (amino acids 90-213); *PAPM2* (5'-GCGGAATTCATGCCGCCTCTTCAGACCCCTGTAA-3') (SEQ ID NO: 199) and *2HMR9* for THAP1-C2 (amino acids 120-213); *PAPM3* (5'-GCGGAATTCATGCACCAGCGGAAAAGGATTCATCAG-3') (SEQ ID NO: 200) and *2HMR9* for THAP1-C3 (amino acids 143-213); *2HMR10* (5'-CCGAATTCAGGATGGTGCAGTCCTGCTCCGCCT-3') (SEQ ID NO: 201) and *2HMR17* (5'-GCGGGATCCCTTGTCATGTGGCTCAGTACAAAGAAATAT-3') (SEQ ID NO: 202) for THAP1-N1 (amino acids 1-90); *2HMR10* and *PAPN2* (5'-CGGGATCCTGTGCGGTCTTGAGCTTCTTTCTGAG-3') (SEQ ID NO: 203) for THAP1-N2 (amino acids 1-166); and *2HMR10* and *PAPN3* (5'-GCGGGATCCGTCGTCTTTCTCTTCTGGAAGTGAAC-3') (SEQ ID NO: 204) for THAP1-N3 (amino acids 1-192).

The PCR fragments, thus obtained, were digested with EcoRI and BamHI, and cloned in frame downstream of the Gal4 Binding Domain (Gal4-BD) in pGBKT7 two-hybrid vector (Clontech) to generate pGBKT7-THAP1-C1, -C2, -C3, -N1, -N2 or -N3, or downstream of the Enhanced Green Fluorescent Protein (EGFP) ORF in pEGFP.C2 vector (Clontech) to generate pEGFP.C2-THAP1-C1, -C2, -C3, -N1, -N2 or -N3.

Two-hybrid interaction between THAP1 mutants and Par4DD was tested by cotransformation of AH109 with pGBKT7-THAP1-C1, -C2, -C3, -N1, -N2 or -N3 and pGADT7-Par4DD and selection of transformants by His and Ade double auxotrophy according to manufacturer's instructions (MATCHMAKER two-hybrid system 3, Clontech). Positive two-hybrid interaction with Par4DD was observed with mutants THAP1-C1, -C2, -C3, -and -N3 but not with mutants THAP1-N1 and -N2, suggesting the Par4 binding site is found between THAP1 residues 143 and 192.

THAP1 mutants were also tested in the *in vitro* THAP1/Par4 interaction assay. *In vitro*-translated THAP1 mutants were generated with the TNT-coupled reticulocyte lysate system (Promega, Madison, WI, USA) using pGBKT7-THAP1-C1, -C2, -C3, -N1, -N2 or -N3 vector as template. 25 µl of each <sup>35</sup>S-labelled THAP1 mutant was incubated with immobilized GST or GST-Par4 protein overnight at 4 °C, in the following binding buffer: 10 mM NaPO<sub>4</sub> pH 8.0, 140 mM NaCl, 3 mM MgCl<sub>2</sub>, 1mM dithiothreitol (DTT), 0.05% NP40, and 0.2 mM phenylmethylsulphonyl fluoride (PMSF), 1 mM Na Vanadate, 50mM β Glycerophosphate, 25 µg/ml

chymotrypsine, 5 µg/ml aprotinin, 10 µg/ml Leupeptin. Beads were then washed 5 times in 1 ml binding buffer. Bound proteins were eluted with 2X Laemmli SDS-PAGE sample buffer, fractionated by 10% SDS-PAGE and visualized by fluorography using Amplify (Amersham Pharmacia Biotech). As expected, THAP1-C1, -C2, -C3, -and -N3 interacted with GST/Par4DD (Figure 4B). In contrast, THAP1-N1 and -N2 failed to interact with GST/Par4DD beads.

Finally, Par4 binding activity of THAP1 mutants was also analyzed by the *in vivo* THAP1/Par4 interaction assay as described in Example 6 using pEF-mycPar4DD and pEGFP.C2-THAP1-C1, -C2, -C3, -N1, -N2 or -N3 expression vectors.

Essentially identical results were obtained with the three THAP1/Par4 interactions assays (Figure 4A). That is, the Par4 binding site was found between residues 143 and 192 of human THAP1. Comparison of this region with the Par4 binding domain of mouse ZIP kinase, another Par4-interacting protein, revealed the existence of a conserved arginine rich-sequence motif (SEQ ID NOs: 205, 263 and 15), that may correspond to the Par4 binding site (Figure 5A). Mutations in this arginine rich-sequence motif were generated by site directed mutagenesis. These two novel THAP1 mutants, THAP1 RR/AA (replacement of residues R171A and R172A) and THAP1ΔQRCRR (deletion of residues 168-172), were generated by two successive rounds of PCR using pEGFP.C2-THAP1 as template and primers 2HMR10 and 2HMR9 together with primers RR/AA-1 (5'-CCGCACAGCAGCGATGCGCTGCTCAAGAACGGCAGCTTG-3') (SEQ ID NO: 206) and RR/AA-2 (5'-CAAGCTGCCGTTCTTGAGCAGCGCATCGCTGCTGTGCGG-3') (SEQ ID NO: 207) for mutant THAP1 RR/AA or primers ΔRR-1 (5'-GCTCAAGACCGCACAGCAAGAACGGCAGCTTG-3') (SEQ ID NO: 208) and ΔRR-2 (5'-CAAGCTGCCGTTCTTGCTGTGCGGTCTTGAGC-3') (SEQ ID NO: 209) for mutant THAP1ΔQRCRR. The resulting PCR fragments were digested with EcoRI and BamHI, and cloned in frame downstream of the Gal4 Binding Domain (Gal4-BD) in pGBKT7 two-hybrid vector (Clontech) to generate pGBKT7-THAP1-RR/AA and -Δ(QRCRR), or downstream of the Enhanced Green Fluorescent Protein (EGFP) ORF in pEGFP.C2 vector (Clontech) to generate pEGFP.C2-THAP1-RR/AA and -Δ(QRCRR). THAP1 RR/AA and THAP1ΔQRCRR THAP1 mutants were then tested in the three THAP1/Par4 interaction assays (two-hybrid assay, *in vitro* THAP1/Par4 interaction assay, *in vivo* THAP1/Par4 interaction assay) as described above for the THAP1-C1, -C2, -C3, -N1, -N2 or -N3 mutants. This analysis revealed that the two mutants were deficient for interaction with Par4 in all three assays (Figure 5B), indicating that the novel arginine-rich sequence motif, we have identified, is a novel Par4 binding motif.

#### EXAMPLE 8

PAR4 is a novel component of PML-NBs that colocalizes with THAP1 *in vivo*

We then wished to determine if PAR4 colocalizes with THAP1 *in vivo* in order to provide further evidence for a physiological interaction between THAP1 and PAR4. We first analyzed Par4 subcellular localization in primary human endothelial cells. Confocal immunofluorescence microscopy using affinity-purified anti-PAR4 antibodies (Sells et al., 1997 ; Guo et al ; 1998) was performed on HUVEC endothelial cells fixed with methanol/acetone, which makes PML-NBs components accessible for antibodies (Sternsdorf et al., 1997). Cells were fixed in methanol for 5 min at -20°C, followed by incubation in cold acetone at -20°C for 30 sec. Permeabilized cells were then blocked with PBS-BSA (PBS with 1% bovine serum albumin) for 10 min and then incubated 2 hr at room temperature with rabbit polyclonal antibodies against human Par4 (1/50, R-334, Santa Cruz Biotechnology, Santa Cruz, CA, USA) and mouse monoclonal antibody anti-PML (mouse IgG1, 1/30, mAb PG-M3 from Dako, Glostrup, Denmark). Cells were then washed three times 5 min at room temperature in PBS-BSA, and incubated for 1 hr with Cy3 (red fluorescence)-conjugated goat anti-rabbit IgG (1/1000, Amersham Pharmacia Biotech) and FITC-labeled goat anti-mouse-IgG (1/40, Zymed Laboratories Inc., San Francisco, CA, USA) secondary antibodies, diluted in PBS-BSA. After extensive washing in PBS, samples were air dried and mounted in Mowiol. Images were collected on a Leica confocal laser scanning microscope. The FITC (green) and Cy3 (red) fluorescence signals were recorded sequentially for identical image fields to avoid cross-talk between the channels. This analysis showed an association of PAR4 immunoreactivity with nuclear dot-like structures, in addition to diffuse nucleoplasmic and cytoplasmic staining. Double immunostaining with anti-PML antibodies, revealed that the PAR4 foci colocalize perfectly with PML-NBs in cell nuclei. Colocalization of Par4 with GFP-THAP1 in PML-NBs was analyzed in transfected HUVEC cells expressing ectopic GFP-THAP1. HUVEC were grown in complete ECGM medium (PromoCell, Heidelberg, Germany), plated on coverslips and transiently transfected with GFP/THAP1 expression construct (pEGFP.C2-THAP1) in RPMI medium using GeneJammer transfection reagent according to manufacturer instructions (Stratagene, La Jolla, CA, USA). Analysis of transfected cells by indirect immunofluorescence microscopy 24h later, with anti-Par4 rabbit antibodies, revealed that all endogenous PAR4 foci colocalize with ectopic GFP-THAP1 in PML-NBs further confirming the association of the THAP1/PAR4 complex with PML-NBs *in vivo*.

30

## EXAMPLE 9

PML recruits the THAP1/PAR4 complex to PML-NBs

Since it has been shown that PML plays a critical role in the assembly of PML-NBs by recruiting other components, we next wanted to determine whether PML plays a role in the recruitment of the THAP1/PAR4 complex to PML-NBs. For this purpose, we made use of the observation that both endogenous PAR4 and ectopic GFP-THAP1 do not accumulate in PML-NBs in human Hela cells. Expression vectors for GFP-THAP1 and HA-PML (or HA-SP100) were cotransfected into these cells and the localization of endogenous PAR4, GFP-THAP1 and HA-PML

35



(or HA-SP100) was analyzed by triple staining confocal microscopy.

Human Hela cells (ATCC) were grown in Dulbecco's Modified Eagle's Medium supplemented with 10% Fetal Calf Serum and 1% Penicillin-streptomycin (all from Life Technologies, Grand Island, NY, USA), plated on coverslips, and transiently transfected with calcium phosphate method using 2 µg pEGFP.C2-THAP1 and pcDNA.3-HA-PML3 or pSG5-HA-Sp100 (a gift from Dr Dejean, Institut Pasteur, Paris, France) plasmid DNA. pcDNA.3-HA-PML3 was constructed by sub-cloning a BglII-BamHI fragment from pGADT7-HA-PML3 into the BamHI site of pcDNA3 expression vector (Invitrogen, San Diego, CA, USA). To generate pGADT7-HA-PML3, PML3 ORF was amplified by PCR, using pACT2-PML3 (a gift from Dr De Thé, Paris, France) as template, with primers  
*PML-1* (5'-GCGGGATCCCTAAATTAGAAAGGGGTGGGGGTAGCC-3') (SEQ ID NO: 210) and  
*PML-2* (5'-GCGGAATTCATGGAGCCTGCACCCGCCCGATC-3') (SEQ ID NO: 211), and cloned into the EcoRI and BamHI sites of pGADT7.

Hela cells transfected with GFP-tagged and HA-tagged expression constructs were allowed to grow for 24 h to 48 h on coverslips. Cells were washed twice with PBS, fixed in methanol for 5 min at -20°C, followed by incubation in cold acetone at -20°C for 30 sec. Permeabilized cells were then blocked with PBS-BSA (PBS with 1% bovine serum albumin) for 10 min and then incubated 2 hr at room temperature with the following primary antibodies diluted in PBS-BSA: rabbit polyclonal antibodies against human Par4 (1/50, R-334, Santa Cruz Biotechnology, Santa Cruz, CA, USA), and mouse monoclonal antibody anti-HA tag (mouse IgG1, 1/1000, mAb 16B12 from BabCO, Richmond, CA, USA). Cells were then washed three times 5 min at room temperature in PBS-BSA, and incubated for 1 hr with Cy3 (red fluorescence)-conjugated goat anti-rabbit IgG (1/1000, Amersham Pharmacia Biotech) and Alexa Fluor-633 (blue fluorescence) goat anti-mouse IgG conjugate (1/100, Molecular Probes, Eugene, OR, USA) secondary antibodies, diluted in PBS-BSA. After extensive washing in PBS, samples were air dried and mounted in Mowiol. Images were collected on a Leica confocal laser scanning microscope. The GFP (green), Cy3 (red) and Alexa 633 (blue) fluorescence signals were recorded sequentially for identical image fields to avoid cross-talk between the channels.

In Hela cells transfected with HA-PML, endogenous PAR4 and GFP-THAP1 were recruited to PML-NBs, whereas in cells transfected with HA-SP100, both PAR4 and GFP-THAP1 exhibited diffuse staining without accumulation in PML-NBs. These findings indicate that recruitment of the THAP1/PAR4 complex to PML-NBs depends on PML but not SP100.

#### EXAMPLE 10

THAP1 is an apoptosis inducing polypeptide

**THAP1 is a novel proapoptotic factor**

Since PML and PML-NBs have been linked to regulation of cell death and PAR4 is a well established pro-apoptotic factor, we examined whether THAP1 can modulate cell survival. Mouse 3T3 cells, which have previously been used to analyze the pro-apoptotic activity of PAR4 (Diaz-Meco et al., 1996; Berra et al., 1997), were transfected with expression vectors for GFP-THAP1, GFP-PAR4 and as a negative control GFP-APSK-1 (APSK-1), a nuclear enzyme unrelated to THAP1 and apoptosis (Girard et al., 1998; Besset et al., 2000). We then determined whether ectopic expression of THAP1 enhances the apoptotic response to serum withdrawal. Transfected cells were deprived of serum for up to twenty four hours and cells with apoptotic nuclei, as revealed by DAPI staining and in situ TUNEL assay, were counted.

10 *Cell death assays:* Mouse 3T3-TO fibroblasts were seeded on coverslips in 12-well plates at 40 to 50% confluency and transiently transfected with GFP or GFP-fusion protein expression vectors using Lipofectamine Plus reagent (Life Technologies) according to supplier's instructions. After 6h at 37°C, the DNA-lipid mixture was removed and the cells were allowed to recover in complete medium for 24 h. Serum starvation of transiently transfected cells was induced by  
15 changing the medium to 0% serum, and the amount of GFP-positive apoptotic cells was assessed 24 h after induction of serum starvation. Cells were fixed in PBS containing 3.7% formaldehyde and permeabilized with 0.1% Triton-X100 as described under immunofluorescence, and apoptosis was scored by *in situ* TUNEL (terminal deoxynucleotidyl transferase-mediated dUTP nick end labeling) and/or DAPI (4,6-Diamidino-2-phenylindole) staining of apoptotic nuclei exhibiting nuclear  
20 condensation. The TUNEL reaction was performed for 1 hr at 37°C using the *in situ* cell death detection kit, TMR red (Roche Diagnostics, Meylan, France). DAPI staining with a final concentration of 0.2 µg/ml was performed for 10 min at room temperature. At least 100 cells were scored for each experimental point using a fluorescence microscope.

Basal levels of apoptosis in the presence of serum ranged from 1-3 %. Twenty four hours  
25 after serum withdrawal, apoptosis was found in 18% of untransfected 3T3 cells and in 3T3 cells overexpressing GFP-APSK-1. Levels of serum withdrawal induced apoptosis were significantly increased to about 70% and 65% in cells overexpressing GFP-PAR4 and GFP-THAP1, respectively (Figure 6A). These results demonstrate that THAP1, similarly to PAR4, is an apoptosis inducing polypeptide.

30 TNFα-induced apoptosis assays were performed by incubating transiently transfected cells in complete medium containing 30 ng/ml of mTNFα (R & D, Minneapolis, MN, USA) for 24 h. Apoptosis was scored as described for serum withdrawal-induced apoptosis. The results are shown in Figure 6B. As shown in Figure 6B, THAP1 induced apoptosis.

#### EXAMPLE 11

##### 35 The THAP domain is essential for THAP1 pro-apoptotic activity

To determine the role of the amino-terminal THAP domain (amino acids 1 to 89) in the functional activity of THAP1, we generated a THAP1 mutant that is deleted of the THAP domain

(THAP1 $\Delta$ THAP). THAP1 $\Delta$ THAP (amino acids 90-213) was amplified by PCR, using pEGFP.C2-THAP1 as template, with primers 2HMR12 (5'-GCGGAATTCAAAGAAGATCTTCTGGAGCCACAGGAAC-3') (SEQ ID NO: 212) and 2HMR9 (5'-CGCGGATCCTGCTGGTACTTCAACTATTTCAAAGTAGTC-3') (SEQ ID NO: 213),  
5 digested with EcoRI and BamHI, and cloned in pGBKT7 and pEGFP-C2 vectors, to generate pGBKT7-THAP1 $\Delta$ THAP and pEGFP.C2-THAP1 $\Delta$ THAP expression vectors. The role of the THAP domain in PML NBs localization, binding to Par4, or pro-apoptotic activity of THAP1 was then analyzed.

To analyze the subcellular localization of THAP1 $\Delta$ THAP, the GFP/ THAP1 $\Delta$ THAP  
10 expression construct was transfected into human primary endothelial cells from umbilical vein (HUVEC, PromoCell, Heidelberg, Germany). HUVEC were grown in complete ECGM medium (PromoCell, Heidelberg, Germany), plated on coverslips and transiently transfected in RPMI medium using GeneJammer transfection reagent according to manufacturer instructions (Stratagene, La Jolla, CA, USA). Transfected cells were allowed to grow for 48 h on coverslips.  
15 Cells were then washed twice with PBS, fixed for 15 min at room temperature in PBS containing 3.7% formaldehyde, and washed again with PBS prior to neutralization with 50mM NH<sub>4</sub>Cl in PBS for 5 min at room temperature. Following one more PBS wash, cells were permeabilized 5 min at room temperature in PBS containing 0.1% Triton-X100, and washed again with PBS. Permeabilized cells were then blocked with PBS-BSA (PBS with 1% bovine serum albumin) for  
20 10' and then incubated 2 hr at room temperature with mouse monoclonal antibody anti-PML (mouse IgG1, 1/30, mAb PG-M3 from Dako, Glostrup, Denmark) diluted in PBS-BSA. Cells were then washed three times 5 min at room temperature in PBS-BSA, and incubated for 1 hr with Cy3 (red fluorescence)-conjugated goat anti-mouse IgG (1/1000, Amersham Pharmacia Biotech) secondary antibodies, diluted in PBS-BSA. After extensive washing in PBS, samples were air dried  
25 and mounted in Mowiol. Images were collected on a Leica confocal laser scanning microscope. The GFP (green) and Cy3 (red) fluorescence signals were recorded sequentially for identical image fields to avoid cross-talk between the channels.

This analysis revealed that GFP- THAP1 $\Delta$ THAP staining exhibits a complete overlap with the staining pattern obtained with antibodies directed against PML, indicating the THAP domain is  
30 not required for THAP1 localization to PML NBs.

To examine the role of the THAP domain in binding to Par4, we performed *in vitro* GST pull down assays. Par4DD, expressed as a GST-tagged fusion protein and immobilized on glutathione sepharose, was incubated with radiolabeled *in vitro* translated THAP1 $\Delta$ THAP. *In vitro*-translated THAP1 $\Delta$ THAP was generated with the TNT-coupled reticulocyte lysate system  
35 (Promega, Madison, WI, USA) using pGBKT7-THAP1 $\Delta$ THAP vector as template. 25  $\mu$ l of <sup>35</sup>S-labelled THAP1 $\Delta$ THAP was incubated with immobilized GST-Par4 or GST proteins overnight at 4 °C, in the following binding buffer : 10 mM NaPO<sub>4</sub> pH 8.0, 140 mM NaCl, 3 mM MgCl<sub>2</sub>, 1mM

dithiothreitol (DTT), 0.05% NP40, and 0.2 mM phenylmethyl sulphonyl fluoride (PMSF), 1 mM Na Vanadate, 50mM  $\beta$  Glycerophosphate, 25  $\mu$ g/ml chymotrypsine, 5  $\mu$ g/ml aprotinin, 10  $\mu$ g/ml Leupeptin. Beads were then washed 5 times in 1 ml binding buffer. Bound proteins were eluted with 2X Laemmli SDS-PAGE sample buffer, fractionated by 10% SDS-PAGE and visualized by  
5 fluorography using Amplify (Amersham Pharmacia Biotech).

This analysis revealed that THAP1 $\Delta$ THAP interacts with GST/Par4DD, indicating that the THAP domain is not involved in THAP1/Par4 interaction (Figure 7A).

To examine the role of the THAP domain in THAP1 pro-apoptotic activity, we performed cell death assays in mouse 3T3 cells. Mouse 3T3-TO fibroblasts were seeded on coverslips in 12-  
10 well plates at 40 to 50% confluency and transiently transfected with GFP-APSK1, GFP-THAP1 or GFP-THAP1 $\Delta$ THAP fusion proteins expression vectors using Lipofectamine Plus reagent (Life Technologies) according to supplier's instructions. After 6h at 37°C, the DNA-lipid mixture was removed and the cells were allowed to recover in complete medium for 24 h. Serum starvation of transiently transfected cells was induced by changing the medium to 0% serum, and the amount of  
15 GFP-positive apoptotic cells was assessed 24 h after induction of serum starvation. Cells were fixed in PBS containing 3.7% formaldehyde and permeabilized with 0.1% Triton-X100 as described under immunofluorescence, and apoptosis was scored by *in situ* TUNEL (terminal deoxynucleotidyl transferase-mediated dUTP nick end labeling) and/or DAPI (4,6-Diamidino-2-phenylindole) staining of apoptotic nuclei exhibiting nuclear condensation. The TUNEL reaction  
20 was performed for 1 hr at 37°C using the *in situ* cell death detection kit, TMR red (Roche Diagnostics, Meylan, France). DAPI staining with a final concentration of 0.2  $\mu$ g/ml was performed for 10 min at room temperature. At least 100 cells were scored for each experimental point using a fluorescence microscope.

Twenty four hours after serum withdrawal, apoptosis was found in 18% of untransfected  
25 3T3 cells and in 3T3 cells overexpressing GFP-APSK-1. Levels of serum withdrawal induced apoptosis were significantly increased to about 70% in cells overexpressing GFP-THAP1. Deletion of the THAP domain abrogated most of this effect since serum-withdrawal-induced apoptosis was reduced to 28 % in cells overexpressing GFP-THAP1 $\Delta$ THAP (Figure 7B). These results indicate that the THAP domain, although not required for THAP1 PML-NBs localization and Par4 binding,  
30 is essential for THAP1 pro-apoptotic activity.

## EXAMPLE 12

### The THAP domain defines a novel family of proteins, the THAP family

To discover novel human proteins homologous to THAP1 and/or containing THAP domains, GenBank non-redundant, human EST and draft human genome databases at the National  
35 Center for Biotechnology Information ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) were searched with both the nucleotide and amino acid sequences of THAP1, using the programs BLASTN, TBLASTN and BLASTP (Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local

alignment search tool. *J Mol Biol* 215: 403-410). This initial step enabled us to identify 12, distinct human THAP-containing, proteins (hTHAP0 to hTHAP11; Figure 8). In the case of the partial length sequences, assembly of overlapping ESTs together with GENESCAN (Burge, C. and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 268: 78-94) and GENEWISE (Jareborg, N., Birney, E. and Durbin, R. (1999). Comparative analysis of noncoding regions of 77 orthologous mouse and human gene pairs. *Genome Res* 9: 815-824) gene predictions on the corresponding genomic DNA clones, was used to define the full length human THAP proteins as well as their corresponding cDNAs and genes. CLUSTALW (Higgins, D. G., Thompson, J. D. and Gibson, T. J. (1996). Using CLUSTAL for multiple sequence alignments. *Methods Enzymol* 266: 383-402) was used to carry out the alignment of the 12 human THAP domains with the DNA binding domain of Drosophila P-element transposase (Lee, C. C., Beall, E. L., and Rio, D. C. (1998) *Embo J.* 17:4166-74), which was colored using the computer program Boxshade ([www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)) (see Figures 9A and 9B). Equivalent approach to the one described above was used in order to identify the mouse, rat, pig, and various other orthologs of the human THAP proteins (Figure 9C). Altogether, the *in silico* and experimental approaches led to the discovery of 12 distinct human members (hTHAP0 to hTHAP11) of the THAP family of pro-apoptotic factors (Figure 8).

#### EXAMPLE 13

##### THAP2 and THAP3 interact with Par-4

To assess whether THAP2 and THAP3 are able to interact with Par-4, yeast two hybrid assays using Par-4 wild type bait (Figure 10B) and *in vitro* GST pull down assays (Figure 10C), were performed as described above (Examples 4 and 5). As shown in Figures 10B and 10C, THAP2 and THAP3 are able to interact with Par-4. A sequence alignment showing the comparison of the THAP domain and the PAR4-binding domain between THAP1, THAP2 and THAP3 is shown in Figure 10A.

#### EXAMPLE 14

##### THAP2 and THAP3 are able to induce apoptosis

Serum-induced or TNF $\alpha$  apoptosis analyses were performed as described above (Example 10) in cells transfected with GFP-APSK1, GFP-THAP2 or GFP-THAP3 expression vectors. Apoptosis was quantified by DAPI staining of apoptotic nuclei 24 hours after serum withdrawal or addition of TNF $\alpha$ . The results are shown in Figure 11A (serum withdrawal) and Figure 11B (TNF $\alpha$ ). These results indicate that, THAP-2 and THAP3 induce apoptosis.

#### EXAMPLE 15

##### Identification of the SLC/CCL21 chemokine-binding domain of human THAP1

To identify the SLC/CCL21 chemokine-binding domain of human THAP1, a series of THAP1 deletion constructs was generated as described in Example 7.

Two-hybrid interaction between THAP1 mutants and chemokine SLC/CCL21 was tested by cotransformation of AH109 with pGADT7-THAP1-C1, -C2, -C3, -N1, -N2 or -N3 and pGBKT7-SLC/CCL21 and selection of transformants by His and Ade double auxotrophy according to manufacturer's instructions (MATCHMAKER two-hybrid system 3, Clontech). pGBKT7-SLC/CCL21 vector was generated by subcloning the BamHI SLC/CCL21 fragment from pGBT9-SLC (see example 1) into the unique BamHI cloning site of vector pGBKT7 (Clontech). Positive two-hybrid interaction with chemokine SLC/CCL21 was observed with mutants THAP1-C1, -C2, -C3, but not with mutants THAP1-N1, -N2 and -N3, suggesting that the SLC/CCL21 chemokine-binding domain of human THAP1 is found between THAP1 residues 143 and 213 (Figure 12).

## EXAMPLE 16

*In vitro* THAP1/chemokine SLC-CCL21 interaction assay

To confirm the interaction observed in yeast two-hybrid system, we performed *in vitro* GST pull down assays. THAP1, expressed as a GST-tagged fusion protein and immobilized on glutathione sepharose, was incubated with radiolabeled *in vitro* translated SLC/CCL21.

To generate the GST-THAP1 expression vector, the full-length coding region of THAP1 (amino acids 1-213) was amplified by PCR from HEVEC cDNA with primers 2HMR8 (5'-CGCGGATCCGTGCAGTCCTGCTCCGCCTACGGC-3') (SEQ ID NO: 214) and 2HMR11 (5'-CCGAATTCTTATGCTGGTACTTCAACTATTTCAAAGTAG-3') (SEQ ID NO: 215), digested with BamHI and EcoRI, and cloned in frame downstream of the Glutathion S-Transferase ORF, between the BamHI and EcoRI sites of the pGEX-2T prokaryotic expression vector (Amersham Pharmacia Biotech, Saclay, France). GST-THAP1 fusion protein encoded by plasmid pGEX-2T-THAP1 and control GST protein encoded by plasmid pGEX-2T, were then expressed in *E. Coli* DH5 $\alpha$  and purified by affinity chromatography with glutathione sepharose according to supplier's instructions (Amersham Pharmacia Biotech). The yield of proteins used in GST pull-down assays was determined by SDS-Polyarylamide Gel Electrophoresis (PAGE) and Coomassie blue staining analysis.

*In vitro*-translated SLC/CCL21 was generated with the TNT-coupled reticulocyte lysate system (Promega, Madison, WI, USA) using as template pGBKT7-SLC/CCL21 vector (see Example 15). 25  $\mu$ l of <sup>35</sup>S-labelled wild-type SLC/CCL21 was incubated with immobilized GST-THAP1 or GST proteins overnight at 4 °C, in the following binding buffer : 10 mM NaPO<sub>4</sub> pH 8.0, 140 mM NaCl, 3 mM MgCl<sub>2</sub>, 1mM dithiothreitol (DTT), 0.05% NP40, and 0.2 mM phenylmethyl sulphonyl fluoride (PMSF), 1 mM Na Vanadate, 50mM  $\beta$  Glycerophosphate, 25  $\mu$ g/ml chymotrypsine, 5  $\mu$ g/ml aprotinin, 10  $\mu$ g/ml Leupeptin. Beads were then washed 5 times in 1 ml binding buffer. Bound proteins were eluted with 2X Laemmli SDS-PAGE sample buffer, fractionated by 10% SDS-PAGE and visualized by fluorography using Amplify (Amersham Pharmacia Biotech). As expected, GST/THAP1 interacted with SLC/CCL21 (Figure 13). In contrast, SLC/CCL21 failed to interact with GST beads.

## EXAMPLE 17

Identification of the THAP1-binding domain of human chemokine SLC/CCL21

To determine the THAP1-binding site on human chemokine SLC/CCL21, a SLC/CCL21 deletion mutant (SLC/CCL21 $\Delta$ COOH) lacking the SLC-specific basic carboxy-terminal extension (amino acids 102-134; GenBank Accession Number NP\_002980) was generated. This SLC/CCL21 $\Delta$ COOH mutant, which retains the CCR7 chemokine receptor binding domain of SLC/CCL21 (amino acids 24-101), was used both in yeast two-hybrid assays with THAP1 bait and in *in vitro* GST-pull down assays with GST-THAP1.

For two-hybrid assays, yeast cells were cotransformed with BD7-THAP1 and AD7-SLC/CCL21 or AD7-SLC/CCL21 $\Delta$ COOH expression vectors. AD7-SLC/CCL21 or AD7-SLC/CCL21 $\Delta$ COOH expression vectors were generated by subcloning BamHI fragment (encoding SLC amino acids 24-134) or BamHI-PstI fragment (encoding SLC amino acids 24-102) from pGKT7-SLC/CCL21 (see example 15) into pGADT7 expression vector (Clontech). Transformants were selected on media lacking histidine and adenine. Figure 13 shows that both the SLC/CCL21 wild type and the SLC/CCL21 $\Delta$ COOH deletion mutants could bind to THAP1. Identical results were obtained by cotransformation of AD7-THAP1 with BD7-SLC/CCL21 or BD7-SLC/CCL21 $\Delta$ COOH.

GST pull down assays, using *in vitro*-translated SLC/CCL21 $\Delta$ COOH, generated with the TNT-coupled reticulocyte lysate system (Promega, Madison, WI, USA) using as template pGBKT7-SLC/CCL21 $\Delta$ COOH, were performed as described in Example 16. Figure 13 shows that both the SLC/CCL21 wild type and the SLC/CCL21 $\Delta$ COOH deletion mutants could bind to THAP1.

## EXAMPLE 18

Preparation of THAP1/Fc Fusion Proteins

This example describes preparation of a fusion protein comprising THAP1 or the SLC/CCL21 chemokine-binding domain of THAP1 fused to an Fc region polypeptide derived from an antibody. An expression vector encoding the THAP1/Fc fusion protein is constructed as follows.

Briefly, the full length coding region of human THAP1 (SEQ ID NO: 3; amino acids -1 to 213) or the SLC/CCL21 chemokine-binding domain of human THAP1 (SEQ ID NO: 3; amino acids -143 to 213) is amplified by PCR. The oligonucleotides employed as 5' primers in the PCR contain an additional sequence that adds a Not I restriction site upstream. The 3' primer includes an additional sequence that encodes the first two amino acids of an Fc polypeptide, and a sequence that adds a Bgl II restriction site downstream of the THAP1 and Fc sequences.

A recombinant vector containing the human THAP1 cDNA is employed as the template in the PCR, which is conducted according to conventional procedures. The amplified DNA is then digested with Not I and Bgl II, and the desired fragments are purified by electrophoresis on an agarose gel.

A DNA fragment encoding the Fc region of a human IgG1 antibody is isolated by digesting a vector containing cloned Fc-encoding DNA with Bgl II and Not I. Bgl II cleaves at a unique Bgl II site introduced near the 5' end of the Fc-encoding sequence, such that the Bgl II site encompasses the codons for amino acids three and four of the Fc polypeptide. Not I cleaves downstream of the Fc-encoding sequence. The nucleotide sequence of cDNA encoding the Fc polypeptide, along with the encoded amino acid sequence, can be found in International Publication No: WO93/10151.

In a three-way ligation, the above-described THAP1 (or SLC/CCL21 chemokine-binding domain of THAP1) -encoding DNA and Fc-encoding DNA are inserted into an expression vector that has been digested with Not I and treated with a phosphatase to minimize recircularization of any vector DNA without an insert. An example of a vector which can be used is pDC406 (described in McMahan et al., EMBO J. 10:2821, 1991), which is a mammalian expression vector that is also capable of replication in *E. coli*.

*E. coli* cells are then transfected with the ligation mixture, and the desired recombinant vectors are isolated. The vectors encode amino acids-1 to 213 of the THAP1 sequence (SEQ ID NO: 3) or amino acids-143 to 213 of the THAP1 sequence of (SEQ ID NO: 3), fused to the N-terminus of the Fc polypeptide. The encoded Fc polypeptide extends from the N-terminal hinge region to the native C-terminus, i.e., is an essentially full-length antibody Fc region.

CV-1/EBNA-1 cells are then transfected with the desired recombinant isolated from *E. coli*. CV-1/EBNA-1 cells (ATCC CRL 10478) can be transfected with the recombinant vectors by conventional procedures. The CVI-EBNA-1 cell line was derived from the African Green Monkey kidney cell line CV-1 (ATCC CCL 70), as described by McMahan et al. (1991). EMBO J. 10:2821. The transfected cells are cultured to allow transient expression of the THAP1/Fc or SLC/CCL21 chemokine-binding domain of THAP1/Fc fusion proteins, which are secreted into the culture medium. The secreted proteins contain the mature form of THAP1 or the SLC/CCL21 chemokine-binding domain of THAP1, fused to the Fc polypeptide. The THAP1/Fc and SLC/CCL21 chemokine-binding domain of THAP1/Fc fusion proteins are believed to form dimers, wherein two such fusion proteins are joined by disulfide bonds that form between the Fc moieties thereof. The THAP1/Fc and SLC/CCL21 chemokine-binding domain of THAP1/Fc fusion proteins can be recovered from the culture medium by affinity chromatography on a Protein A-bearing chromatography column.

#### EXAMPLE 19

##### The THAP domain defines a family of nuclear factors

To determine the subcellular localization of the different human THAP proteins, a series of GFP-THAP expression constructs were transfected into primary human endothelial cells. In agreement with the possible functions of THAP proteins as DNA-binding factors, we found that all the human THAP proteins analyzed (THAP0, 1, 2, 3, 6, 7, 8, 10, 11) localize preferentially to the cell nucleus (Figure 14). In addition to their diffuse nuclear localization, some of the THAP



proteins also exhibited association with distinct subnuclear structures: the nucleolus for THAP2 and THAP3, and punctuate nuclear bodies for THAP7, THAP8 and THAP11. Indirect immunofluorescence microscopy with anti-PML antibodies revealed that the THAP8 and THAP11 nuclear bodies colocalize with PML-NBs. Although the THAP7 nuclear bodies often appeared in close association with the PML-NBs, they never colocalized.

Analysis of the subcellular localization of the GFP-THAP fusion proteins was performed as described above (Example 3). The GFP-THAP constructs were generated as follows: the human THAP0 coding region was amplified by PCR from Hevec cDNA with primers *THAP0-1* (5'-GCCGAATTCATGCCGAAGCTTCTGCGCTGCCCC-3') (SEQ ID NO: 216) and *THAP0-2* (5'-CGCGGATCCTTAGGTTATTTCCACAGTTTCGGAATTATC-3') (SEQ ID NO: 217), digested with EcoRI and BamHI, and cloned in the same sites of the pEGFP-C2 vector, to generate pEGFPC2-THAP0; the coding region of human THAP2, 3, 7, 6 and 8 were amplified by PCR respectively from Image clone No: 3606376 with primers *THAP2-1* (5'-GCGCTGCAGCAAGCTAAATTTAAATGAAGGTACTCTTGG-3') (SEQ ID NO: 218) and *THAP2-2* (5'-GCGAGATCTGGGAAATGCCGACCAATTGCGCTGCG-3') (SEQ ID NO: 219) digested with BglII and PstI, from Image clone No: 4813302 and No: 3633743 with primers *THAP3-1* (5'-AGAGGATCCTTAGCTCTGCTGCTCTGGCCCAAGTC-3') (SEQ ID NO: 220) *THAP3-2* (5'-AGAGAATTCATGCCGAAGTCGTGCGCGGCCCG-3') (SEQ ID NO: 221) and primers *THAP7-1* (5'-GCGGAATTCATGCCGCGTCACTGCTCCGCCGC-3') (SEQ ID NO: 222) *THAP7-2* (5'-GCGGGATCCTCAGGCCATGCTGCTGCTCAGCTGC-3') (SEQ ID NO: 223), digested with EcoRI and BamHI, from Image clone No: 757753 with primers *THAP6-1* (5'-GCGAGATCTCGATGGTGAAATGCTGCTCCGCCATTGGA-3') (SEQ ID NO: 224) and *THAP6-2* (5'-GCGGGATCCTCATGAAATATAGTCCTGTTCTATGCTCTC-3') (SEQ ID NO: 225) digested with BglII and BamHI, and from Image clone No: 4819178 with primers *THAP8-1* (5'-GCGAGATCTCGATGCCCAAGTACTGCAGGGCGCCG-3') (SEQ ID NO: 226) and *THAP8-2* (5'-GCGGAATTCCTTATGCACTGGGGATCCGAGTGTCCAGG-3') (SEQ ID NO: 227), digested with BglII and EcoRI and cloned in frame downstream of the Enhanced Green Fluorescent Protein (EGFP) ORF in pEGFPC2 vector (Clontech) digested with the same enzymes to generate pEGFPC2-THAP2, -THAP3, -THAP7, -THAP6 and -THAP8; the human THAP10 and THAP11 coding region were amplified by PCR from Hela cDNA respectively with primers *THAP10-1* (5'-GCGGAATTCATGCCGGCCCGTTGTGTGGCCGC-3') (SEQ ID NO: 228) *THAP10-2* (5'-GCGGGATCCTTAACATGTTTCTTCTTTACCTGTACAGC-3') (SEQ ID NO: 229) digested with EcoRI and BamHI, and with primers *THAP11-1* (5'-GCGAGATCTCGATGCCTGGCTTTACGTGCTGCGTGC-3') (SEQ ID NO: 230) and *THAP11-2* (5'-GCGGAATTCCTCACATTCCGTGCTTCTTGCGGATGAC-3') (SEQ ID NO: 231), digested with BglII and EcoRI, cloned in the same sites of the pEGFP-C2 vector, to generate pEGFPC2-THAP10 and -THAP11.

## EXAMPLE 20

The THAP domain shares structural similarities with  
the DNA-binding domain of nuclear hormone receptors

In an effort to model the three-dimensional structure of the THAP domain, we searched the PDB crystallographic database. As sequence homology detection is more sensitive and selective when aided by secondary structure information, structural homologs of the THAP domain of human THAP1 were searched using the SeqFold threading program (Olszewski et al. (1999) Theor. Chem. Acc. 101, 57-61) which combines sequence and secondary structure alignment. The crystallographic structure of the thyroid hormone receptor  $\beta$  DBD (PDB code: 2NLL) gave the best score of the search and we used the resulting structural alignment, displayed in Figure 15A, to derive a homology-based model of the THAP domain from human THAP1 (Figure 15B). Note that the distribution of Cys residues in the THAP domain does not fully match that of the thyroid hormone receptor  $\beta$  DBD (Figure 15A) and hence cannot allow the formation of the two characteristic 'C4-type' Zn-fingers (red color-coding in Figure 15A). However, a network of stacking interactions between aromatic/hydrophobic residues or aliphatic parts of lysine side-chains ensures the stability of the structure of the THAP domain (cyan color-coding in Figures 15A and 15B). Interestingly the same threading method applied independently to the *Drosophila* P-element transposase DBD identified the crystallographic structure of the glucocorticoid receptor DBD (PDB code: 1GLU) as giving the best score. In the same way, we used the resulting structural alignment, displayed in Figure 15D, to build a model of the transposase DBD (Figure 15C). Note the presence of an hydrophobic core equivalent to that of the THAP domain (cyan color-coding in Figures 15C and 15D). All the DNA-binding domains of the nuclear receptors fold into a typical pattern which is mainly based on two interacting  $\alpha$ -helices, the first one inserting into the target DNA major groove. Our threading and modeling results indicate that the THAP domain and the *D. melanogaster* P-element transposase DBD likely share a common topology which is similar to that of the DBD of nuclear receptors.

Molecular modeling was performed using the InsightII, SeqFold, Homology and Discover modules from the Accelrys (San Diego, CA) molecular modeling software (version 98), run on a Silicon Graphics O2 workstation. Optimal secondary structure prediction of the query protein domains was ensured by the DSC method within SeqFold. The threading-derived secondary structure alignments was used as input for homology-modeling, which was performed according to a previously described protocol (Manival et al. (2001) Nucleic Acids Res 29 :2223-2233). The validity of the models was checked both by Ramachandran analysis and folding consistency verification as previously reported (Manival et al. (2001) Nucleic Acids Res 29 :2223-2233).

## EXAMPLE 21

Homodimerization domain of human THAP1

To identify the sequences mediating homodimerization of THAP1, a series of THAP1 deletion constructs was generated as described in Example 7.

Two-hybrid interaction between THAP1 mutants and THAP1 wild type was tested by cotransformation of AH109 with pGADT7-THAP1-C1, -C2, -C3, -N1, -N2 or -N3 and pGBKT7-THAP1 wild-type and selection of transformants by His and Ade double auxotrophy according to manufacturer's instructions (MATCHMAKER two-hybrid system 3, Clontech). Positive two-hybrid interaction with THAP1 wild type was observed with mutants THAP1-C1, -C2, -C3, -and -N3 but not with mutants THAP1-N1 and -N2, suggesting the THAP1 homodimerization domain is found between THAP1 residues 143 and 192 (Figure 16A).

To confirm the results obtained in yeast, THAP1 mutants were also tested in *in vitro* GST pull down assays. Wild type THAP1 expressed as a GST-tagged fusion protein and immobilized on glutathione sepharose (as described in example 16), was incubated with radiolabeled *in vitro* translated THAP1 mutants. *In vitro*-translated THAP1 mutants were generated with the TNT-coupled reticulocyte lysate system (Promega, Madison, WI, USA) using pGADT7-THAP1-C1, -C2, -C3, -N1, -N2 or -N3 vector as template. 25 µl of each <sup>35</sup>S-labelled THAP1 mutant was incubated with immobilized GST or GST-THAP1 wild-type protein overnight at 4 °C, in the following binding buffer: 10 mM NaPO<sub>4</sub> pH 8.0, 140 mM NaCl, 3 mM MgCl<sub>2</sub>, 1mM dithiothreitol (DTT), 0.05% NP40, and 0.2 mM phenylmethyl sulphonyl fluoride (PMSF), 1 mM Na Vanadate, 50mM β Glycerophosphate, 25 µg/ml chymotrypsine, 5 µg/ml aprotinin, 10 µg/ml Leupeptin. Beads were then washed 5 times in 1 ml binding buffer. Bound proteins were eluted with 2X Laemmli SDS-PAGE sample buffer, fractionated by 10% SDS-PAGE and visualized by fluorography using Amplify (Amersham Pharmacia Biotech). As expected, THAP1-C1, -C2, -C3, -and -N3 interacted with GST/THAP1 (Figure 16B). In contrast, THAP1-N1 and -N2 failed to interact with GST/THAP1 beads. Therefore, essentially identical results were obtained with the two THAP1/THAP1 interactions assays: the THAP1 homodimerization domain of THAP1 is found between residues 143 and 192 of human THAP1.

## EXAMPLE 22

### Alternatively spliced isoform of human THAP1

The two distinct THAP1 cDNAs, THAP1a and THAP1b have been discovered (Figure 17A). These splice variants, were amplified by PCR from HEVEC cDNA with primers 2HMR10 (5'-CCGAATTCAGGATGGTGCAGTCCTGCTCCGCCT-3') (SEQ ID NO: 232) and 2HMR9 (5'-CGCGGATCCTGCTGGTACTTCAACTATTTCAAAGTAGTC-3') (SEQ ID NO: 233), digested with EcoRI and BamHI, and cloned in frame upstream of the Enhanced Green Fluorescent Protein (EGFP) ORF in pEGFP.N3 vector (Clontech) to generate pEGFP.N3-THAP1a and pEGFP-THAP1b. DNA sequencing revealed that THAP1b cDNA isoform lacks exon 2 (nucleotides 273-468) of the human THAP1 gene (Figure 17B). This alternatively spliced isoform of human THAP1 (~ 2 kb mRNA) was also observed in many other tissues by Northern blot analysis (see Figure 2).

The THAP1a/GFP and THAP1b/GFP expression constructs were then transfected into COS 7 cells (ATCC) and expression of the fusion proteins was analyzed by western blotting with anti-GFP antibodies. The results are shown in Figure 17C which demonstrates that the second isoform of human THAP1 (THAP1b) encodes a truncated THAP1 protein (THAP1 C3) lacking a substantial portion of the amino terminus (amino acids 1-142 of SEQ ID NO: 3).

## EXAMPLE 23

High throughput screening assay for modulators of THAP familyPolypeptide pro-apoptotic activity

A high throughput screening assay for molecules that abrogate or stimulate THAP-family polypeptide proapoptotic activity was developed, based on serum-withdrawal induced apoptosis in a 3T3 cell line with tetracycline-regulated expression of a THAP family polypeptide.

In a preferred example, the THAP1 cDNA with an in-frame myc tag sequence, was amplified by PCR using pGBKT7-THAP1 as a template with primers *myc.BD7* (5'-GCGCTCTAGAGCCATCATGGAGGAGCAGAAGCTGATC-3') (SEQ ID NO: 234) and *2HMR15* (5'-GCGCTCTAGATTATGCTGGTACTTCAACTATTTCAAAGTAG-3') (SEQ ID NO: 235), and cloned downstream of a tetracycline regulated promoter in plasmid vector pTRE (Clontech, Palo Alto, CA), using *Xba I* restriction site, to generate plasmid pTRE-mycTHAP1. To establish 3T3-TO-mycTHAP1 stable cell lines, mouse 3T3-TO fibroblasts (Clontech) were seeded at 40 to 50% confluency and co-transfected with the pREP4 plasmid (Invitrogen), which contains a hygromycin B resistance gene, and the mycTHAP1 expression vector (pTRE-mycTHAP1) at 1:10 ratio, using Lipofectamine Plus reagent (Life Technologies) according to supplier's instructions. Transfected cells were selected in medium containing hygromycin B (250 U/ml; Calbiochem) and tetracycline (2 ug/ml; Sigma). Several resistant colonies were picked and analyzed for the expression of mycTHAP1 by indirect immunofluorescence using anti-myc epitope monoclonal antibody (mouse IgG1, 1/200, Clontech). A stable 3T3-TO cell line expressing mycTHAP1 (3T3-TO-mycTHAP1) was selected and grown in Dulbecco's Modified Eagle's Medium supplemented with 10% Fetal Calf Serum, 1% Penicillin-streptomycin (all from Life Technologies, Grand Island, NY, USA) and tetracycline (2 ug/ml; Sigma). Induction of THAP1 expression into this 3T3-TO-mycTHAP1 cell line was obtained 48 h after removal of tetracycline in the complete medium.

A drug screening assay using the 3T3-TO-mycTHAP1 cell line can be carried out as follows. 3T3-TO-mycTHAP1 cells are plated in 96- or 384-wells microplates and THAP1 expression is induced by removal of tetracycline in the complete medium. 48 h later, the apoptotic response to serum withdrawal is assayed in the presence of a test compound, allowing the identification of test compounds that either enhance or inhibit the ability of THAP1 polypeptide to induce apoptosis. Serum starvation of 3T3-TO-mycTHAP1 cells is induced by changing the medium to 0% serum, and the amount of cells with apoptotic nuclei is assessed 24 h after induction of serum starvation by TUNEL labeling in 96- or 384-wells microplates. Cells are fixed in PBS

containing 3.7% formaldehyde and permeabilized with 0.1% Triton-X100, and apoptosis is scored by *in situ* TUNEL (terminal deoxynucleotidyl transferase-mediated dUTP nick end labeling) staining of apoptotic nuclei for 1 hr at 37°C using the *in situ* cell death detection kit, TMR red (Roche Diagnostics, Meylan, France). The intensity of TMR red fluorescence in each well is then  
5 quantified to identify test compounds that modify the fluorescence signal and thus either enhance or inhibit THAP1 pro-apoptotic activity.

## EXAMPLE 24

High throughput two-hybrid screening assay for drugs that modulate THAP-family  
polypeptide/THAP-family target protein interaction

10 To identify drugs that modulate THAP1/Par4 or THAP1/SLC interactions, a two-hybrid based high throughput screening assay can be used.

As described in Example 17, AH109 yeast cells (Clontech) cotransformed with plasmids pGBKT7-THAP1 and pGADT7-Par4 or pGADT7-SLC can be grown in 384-well plates in selective media lacking histidine and adenine, according to manufacturer's instructions  
15 (MATCHMAKER two-hybrid system 3, Clontech).

Growth of the transformants on media lacking histidine and adenine is absolutely dependent on the THAP1/Par4 or THAP1/SLC two-hybrid interaction and drugs that disrupt THAP1/Par4 or THAP1/SLC binding will therefore inhibit yeast cell growth.

20 Small molecules (5 mg/ml in DMSO; Chembridge) are added by using plastic 384-pin arrays (Genetix). The plates are incubated for 4 to 5 days at 30 °C, and small molecules which inhibit the growth of yeast cells by disrupting THAP1/Par4 or THAP1/SLC two-hybrid interaction are selected for further analysis.

## EXAMPLE 25

High throughput in vitro assay to identify inhibitors of  
THAP-family polypeptide/THAP-family protein target interaction

25 To identify small molecule modulators of THAP function, a high-throughput screen based on fluorescence polarization (FP) is used to monitor the displacement of a fluorescently labelled THAP1 protein from a recombinant glutathione-S-transferase (GST)-THAP binding domain of Par4 (Par4DD) fusion protein or a recombinant GST-SLC/CCL21 fusion protein.

30 Assays are carried out essentially as in Degterev et al, Nature Cell Biol. 3: 173-182 (2001) and Dandliker et al, Methods Enzymol. 74: 3-28 (1981). The assay can be calibrated by titrating a THAP1 peptide labelled with Oregon Green with increasing amounts of GST-Par4DD or GST-SLC/CCL21 proteins. Binding of the peptide is accompanied by an increase in polarization (mP, millipolarization).

35 THAP 1 and PAR4 polypeptides and GST-fusions can be produced as previously described. The THAP1 peptide was expressed and purified using a QIAexpressionist kit (Qiagen) according to the manufacturer's instructions. Briefly, the entire THAP1 coding sequence was

amplified by PCR using pGBKT7-THAP1 as a template with primers 2HMR8 (5'-CGCGGATCCGTGCAGTCCTGCTCCGCCTACGGC-3') (SEQ ID NO: 236) and 2HMR9 (5'-CGCGGATCCTGCTGGTACTTCAACTATTTCAAAGTAGTC-3') (SEQ ID NO: 237), and cloned into the *Bam*HI site of pQE30 vector (Qiagen). The resulting pQE30-HisTHAP1 plasmid was transformed in *E.coli* strain M15 (Qiagen). 6xHis-tagged-THAP1 protein was purified from inclusion bodies on a Ni-Agarose column (Qiagen) under denaturing conditions, and the eluate was used for in vitro interaction assays. To produce GST-Par4DD fusion protein, Par4DD (amino acids 250-342) was amplified by PCR with primers *Par4.10* (5'-GCCGGATCCGGGTTCCCTAGATATAACAGGGGATGCAA-3') (SEQ ID NO: 238) and *Par4.5* (5'-GCGGGATCCCTCTACCTGGTCAGCTGACCCACAAC-3') (SEQ ID NO: 239), and cloned in frame downstream of the Glutathione S-Transferase (GST) ORF, into the *Bam*HI site of the pGEX-2T prokaryotic expression vector (Amersham Pharmacia Biotech, Saclay, France). Similarly, to produce GST-SLC/CCL21 fusion protein, the mature form of human SLC/CCL21 (amino acids 24-134) was amplified by PCR with primers hSLCbam.5' (5'-GCGGGATCCAGTGATGGAGGGGCTCAGGACTGTTG-3') (SEQ ID NO: 240) and hSLCbam.3' (5'-GCGGGATCCCTATGGCCCTTTAGGGGTCTGTGACC-3') (SEQ ID NO: 241), digested with *Bam*HI and inserted into the *Bam*HI cloning site of the pGEX-2T vector. GST-Par4DD (amino acids 250-342) and GST-SLC/CCL21 (amino acids 24-134) fusion proteins were expressed in *E.Coli* DH5 $\alpha$  (supE44, DELTA $\text{lacU169}$  (80 $\text{lacZ}$ deltaM15), hsdR17, recA1, endA1, gyrA96, thi1, relA 1) and purified by affinity chromatography with glutathione sepharose according to supplier's instructions (Amersham Pharmacia Biotech).

For screening small molecules, THAP1 peptide is labelled with succinimidyl Oregon Green (Molecular Probes, Eugene, Oregon) and purified by HPLC. 33 nM labelled THAP1 peptide, 2 $\mu$ M GST-Par4DD or GST-SLC/CCL21 protein, 0.1% bovine gamma-globulin (Sigma) and 1 mM dithiothreitol mixed with PBS, pH 7.2 (Gibco), are added to 384-well black plates (Lab Systems) with Multidrop (Lab Systems). Small molecules (5 mg/ml in DMSO; Chembridge) are transferred by using plastic 384-pin arrays (Genetix). The plates are incubated for 1-2 hours at 25 °C, and FP values are determined with an Analyst plate reader (LJL Biosystems).

#### EXAMPLE 26

##### High throughput chip assay to identify inhibitors of THAP-family polypeptide/THAP-family protein target interaction

A chip based binding assay Degterev et al, (2001) Nature Cell Biol. 3: 173-182 using unlabelled THAP and THAP-family target protein may be used to identify molecules capable of interfering with THAP-family and THAP-family target interactions, providing high sensitivity and avoiding potential interference from label moieties. In this example, the THAP1 binding domain of Par4 protein (Par4DD) or SLC/CCL21 is covalently attached to a surface-enhanced laser

desorption/ionization (SELDI) chip, and binding of unlabelled THAP1 protein to immobilized protein in the presence of a test compound is monitored by mass spectrometry.

Recombinant THAP1 protein, GST-Par4DD and GST-SLC/CCL21 fusion proteins are prepared as described in Example 25. Purified recombinant GST-Par4DD or GST-SLC/CCL21 protein is coupled through its primary amine to SELDI chip surfaces derivatized with cabonyldiimidazole (Ciphergen). THAP1 protein is incubated in a total volume of 1 µl for 12 hours at 4 °C in a humidified chamber to allow binding to each spot of the SELDI chip, then washed with alternating high-pH and low-pH buffers (0.1M sodium acetate containing 0.5M NaCl, followed by 0.01 M HEPES, pH 7.3). The samples are embedded in an alpha-cyano-4-hydroxycinnamic acid matrix and analyzed for mass by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry. Averages of 100 laser shots at a constant setting are collected over 20 spots in each sample.

#### EXAMPLE 27

##### High throughput cell assay to identify inhibitors of

##### THAP-family polypeptide/THAP-family protein target interaction

A fluorescence resonance energy transfer (FRET) assay is carried out between THAP-1 and PAR4 or SLC/CCL21 proteins fused with fluorescent proteins. Assays can be carried out as in Majhan et al, Nature Biotechnology 16: 547-552 (1998) and Degterev et al, Nature Cell Biol. 3: 173-182 (2001).

THAP-1 protein is fused to cyan fluorescent protein (CFP) and PAR4 or SLC/CCL21 protein is fused to yellow fluorescent protein (YFP). Vectors containing THAP-family and THAP-family target proteins can be constructed essentially as in Majhan et al (1998). A THAP-1-CFP expression vector is generated by subcloning a THAP-1 cDNA into the pECFP-N1 vector (Clontech). PAR4-YFP or SLC/CCL21-CYP expression vectors are generated by subcloning a PAR4 or a SLC/CCL21 cDNA into the pEYFP-N1 vector (Clontech).

Vectors are cotransfected to HEK-293 cells and cells are treated with test compounds. HEK-293 cells are transfected with THAP-1-CFP and PAR4-YFP or SLC/CCL21-YFP expression vectors using Lipofect AMINE Plus (Gibco) or TransLT-1 (PanVera). 24 hours later cells are treated with test compounds and incubated for various time periods, preferably up to 48 hours. Cells are harvested in PBS, optionally supplemented with test compound, and fluorescence is determined with a C-60 fluorimeter (PTI) or a Wallac plate reader. Fluorescence in the samples separately expressing THAP-1-CFP and PAR4-YFP or SLC/CCL21-YFP is added together and used to estimate the FRET value in the absence of THAP-1/PAR4 or THAP1/SLC/CCL21 binding.

The extent of FRET between CFP and YFP is determined as the ratio between the fluorescence at 527nm and that at 475nm after excitation at 433nm. The cotransfection of THAP-1 protein and PAR4 or SLC/CCL21 protein results in an increase of FRET ratio over a reference FRET ratio of 1.0 (determined using samples expressing the proteins separately). A change in the

FRET ratio upon treatment with a test compound (over that observed after cotransfection in the absence of a test compound) indicates a compound capable of modulating the interaction of the THAP-1 protein and the PAR4 or the SLC/CCL21 protein.

## EXAMPLE 28

5                    In vitro assay to identify THAP-family polypeptide DNA targets

DNA binding specificity of THAP1 was determined using a random oligonucleotide selection method allowing unbiased analysis of binding sites selected by the THAP domain of the THAP1 protein from a random pool of possible sites. The method was carried out essentially as described in Bouvet (2001) *Methods Mol Biol.* 148:603-10. Also, see Pollack and Treisman (1990) 10 *Nuc. Acid Res.* 18:6197-6204; Blackwell and Weintraub, (1990) *Science* 250: 1104-1110; Ko and Engel, (1993) *Mol. Cell. Biol.* 13:4011-4022; Merika and Orkin, (1993) *Mol. Cell. Biol.* 13: 3999-4010; and Krueger and Morimoto, (1994) *Mol. Cell. Biol.* 14:7592-7603).

**Recombinant THAP domain expression and purification**

A cDNA fragment encoding the THAP domain of human THAP-1 (amino acids 1-90, SEQ 15 ID NO: 3) was cloned by PCR using as a template pGADT7-THAP-1 (see Example 4) with the following primers 5'-GCGCATATGGTGCAGTCCTGCTCCGCCTACGGC-3' (SEQ ID NO: 242) and 5'-GCGCTCGAGTTTCTTGTCATGTGGCTCAGTACAAAG-3' (SEQ ID NO: 243). The PCR product was cloned as a NdeI-XhoI fragment into pET-21c prokaryotic expression vector (Novagen) in frame with a sequence encoding a carboxy terminal His tag, to generate pET-21c- 20 THAP.

For the expression of THAP-His6, pET-21c-THAP was transformed into Escherichia coli strain BL-21 pLysS. Bacteria were grown at 37°C to an optical density at 600nm of 0.6 and expression of the protein was induced by adding isopropyl-β-D-thiogalactoside (Sigma) at a final concentration of 1mM and incubation was continued for 4 hours.

25                    The cells were collected by centrifugation and resuspended in ice cold of buffer A (50 mM sodium-phosphate pH 7.5, 300mM NaCl, 0.1% β-mercaptoethanol, 10 mM Imidazole). Cells were lysed by sonication and the lysate was cleared by centrifugation at 12000g for 45 min. The supernatant was loaded onto a Ni-NTA agarose column (Quiagen) equilibrated in buffer A. After washing with buffer A and Buffer A with 40 mM Imidazole, the protein was eluted with buffer B 30 (same as A with 0.05%β-mercaptoethanol and 250 mM Imidazole).

Fractions containing THAP-His6 were pooled and applied to a Superdex 75 gel filtration column equilibrated in Buffer C (Tris-HCl 50mM pH 7.5, 150 mM NaCl, 1 mM DTT). Fractions containing the THAP-His6 protein were pooled, concentrated by withn YM-3 Amicon filter devices and stored at 4°C or frozen at -80°C in buffer C containing 20% glycerol. The purity of the sample 35 was assessed by SDS-Polyarylamide Gel Electrophoresis (PAGE) and Coomassie blue staining analysis. The structural integrity of the protein preparation was checked by ESI mass spectrometry



and Peptide mass mapping using a MALDI-TOF Mass spectrometer. The protein concentration was determined with Bradford Protein Assay.

#### Random Oligonucleotide Selection

According to the SELEX protocol described in Bouvet (2001) *Methods Mol Biol.* 148:603-609, a 62 bp oligonucleotide having sequences as follows was synthesized:  
5'-TGGGCACTATTTATATCAAC-N25-AATGTCGTTGGTGGCCC-3' (SEQ ID NO: 244)  
where N is any nucleotide, and primers complementary to each end. Primer P is: 5'-ACCGCAAGCTTGGGCACTATTTATATCAAC-3' (SEQ ID NO: 245), and primer R is 5'-GGTCTAGAGGGCCACCAACGCATT-3' (SEQ ID NO: 246).  
The 62-mer oligonucleotide is made double stranded by PCR using the P and R primers generating a 80 bp random pool.

About 250 ng of THAP-His6 was incubated with Ni-NTA magnetic beads in NT2 buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.05% NP-40) for 30 min at 4°C on a roller. The beads were washed 2 times with 500 µl of NT2 buffer to remove unbound protein. The immobilized THAP-His6 was incubated with the random pool of double stranded 80 bp DNA (2 to 5 µg) in 100 µl of Binding buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.05% NP-40, 0.5 mM EDTA, 100 µg/ml BSA, and 20 to 50 µg of poly(dI-dC)) for 10 minutes at room temperature. The beads were then washed 6 times with 500 µl of NT2 buffer. The protein/DNA complex were then subjected to extraction with phenol/chloroform and precipitation with ethanol using 10 µg of glycogen as a carrier. About one fifth of the recovered DNA was then amplified by 15 to 20 cycles of PCR and used for the next round of selection. After 8 rounds of selection, the NaCl concentration was progressively increased to 150mM.

After 12 rounds of selection by THAP-His6, pools of amplified oligonucleotides were digested with Xba I and Hind III and cloned into pBluescript II KS - (Stratagene) and individual clones were sequenced using Big Dye terminator Kit (Applied Biosystem).

The results of the sequence analysis show that the THAP domain of human THAP1 is a site-specific DNA binding domain. Two consensus sequences were deduced from the alignment of two sets of nucleotide sequences obtained from the above SELEX procedure (each set containing 9 nucleic acid sequences). In particular, it was found that the THAP domain recognizes GGGCAA or TGGCAA DNA target sequences preferentially organized as direct repeats with 5 nucleotide spacing (DR-5 motifs). The consensus sequence being GGGCAAnnnnnTGGCAA (SEQ ID NO: 149). Additionally, THAP recognizes everted repeats with 11 nucleotide spacing (ER-11 motifs) having a consensus sequence of TTGCCAnnnnnnnnnnGGGCAA (SEQ ID NO: 159). Although GGGCAA and TGGCAA sequences constitute the preferential THAP domain DNA binding sites, GGGCAT, GGGCAG and TGGCAG sequences are also DNA target sequences recognized by the THAP domain.

#### EXAMPLE 28B

The THAP domain is a zinc-dependent sequence specific DNA-binding domain

To confirm that the THAP domain is a novel sequence specific DNA-binding domain, electrophoretic mobility shift assays (EMSA) were carried out using wild-type or mutant THAP domain responsive elements (THRE) determined by SELEX (see Example 28 and Figures 18 and 24). Double-stranded probes used in EMSA experiments were purified on 12% polyacrylamide gels, <sup>32</sup>P end labeled with T4 polynucleotide kinase and quantified by Cerenkov counting. About 20 ng of purified THAP domain from human THAP1 (prepared as described in Example 28) was incubated with 30000 cpm of the appropriate probe (about 2 ng). Binding reactions were carried out for 10 minutes at room temperature in 20 µl binding buffer (20 mM Tris-HCl pH 7.5, 100 mM KCl, 0.1% NP-40, 100 µg/ml BSA, 2.5 mM DTT, 5% glycerol, 200 ng poly (dI-dC)). Electrophoresis was performed on 8% (29:1) polyacrylamide gels containing 5% glycerol. Gels were run in 0.25X TBE at 150V and 4°C, dried and exposed on a phosphorimager screen (Molecular Dynamics). Sequences of wild type and mutant THRE oligonucleotides used in EMSA experiments were as follow (mutations are indicated in bold): wild-type probe 3, 5'-AGCAAGTAAGGGCAA**AA**CTACTTCAT-3' (SEQ ID NO: 313); mutant probe 3mut1, 5'-AGCAAGTAATTTCAA**AA**CTACTTCAT-3' (SEQ ID NO: 314); mutant probe 3mut3, 5'-AGCAAGTAAGGTCAA**AA**CTACTTCAT-3' (SEQ ID NO: 319); mutant probe 3mut4, 5'-AGCAAGTAAGTGCAA**AA**CTACTTCAT-3' (SEQ ID NO: 320); mutant probe 3mut14, 5'-AGCAAGTAAGGGCCA**AA**CTACTTCAT-3' (SEQ ID NO: 321); mutant probe 3mut5, 5'-AGCAAGTAAGGGAAA**AA**CTACTTCAT-3' (SEQ ID NO: 322).

These EMSA assays revealed that the THAP domain recognizes wild-type (probe 3) but not mutant THRE oligonucleotides (probes 3mut1, 3mut3, 3mut4, 3mut14, 3mut5) (Figure 25A). For competition experiments, 50-, 150-, and 250-fold molar excess of unlabelled wild-type (THRE competitor, probe 3) or mutant (non-specific competitor, probe 3mut1) THRE oligonucleotides were added to the reaction mixture just before the addition of the probe. This analysis revealed that the DNA-binding activity of the THAP domain is abrogated by increasing amounts of the THRE competitor but not affected by the non-specific competitor (Figure 25B). Together, these experiments demonstrated that the THAP domain is a novel sequence-specific DNA-binding domain.

Since the THAP domain is characterized by a C2-CH conserved motif that may function as a Zn-binding site, we then determined whether DNA-binding activity of the THAP domain is Zn-dependent. For metal chelation experiments, EDTA (5 mM or 50 mM) or 1,10 phenanthroline (Sigma, 1mM or 5mM in methanol vehicle) were preincubated with the THAP domain in binding buffer for 20 minutes at room temperature, before the EMSA assay (Figure 26A). To reconstitute DNA-binding activity of the THAP domain in the presence of 1,10 phenanthroline (+ Phe, 5mM), Zn or Mg, as indicated, were added at 100 or 500 µM final concentration in binding buffer (Figure 26B). Reactions were allowed to equilibrate for 10 minutes at room temperature before the

addition of the EMSA THRE probe (probe 3). These analyses revealed that the DNA-binding activity of the THAP domain is abrogated by the metal-chelator 1,10 phenanthroline (Figure 26A) but specifically restored by the addition of Zinc (Figure 26B), indicating that the THAP domain is a novel zinc-dependent sequence-specific DNA-binding domain.

## EXAMPLE 29

High throughput in vitro assay to identify inhibitors ofTHAP-family polypeptide or THAP-family interactions with nonspecific DNA targets

High throughput assays for the detection and quantification of THAP1-nonspecific DNA binding is carried out using a scintillation proximity assay. Materials are available from Amersham (Piscataway, NJ) and assays can be carried out according to Gal S. et al, 6<sup>th</sup> Ann. Conf. Soc. Biomol. Screening, 6-9 Sept 2000, Vancouver, B.C.).

Random double stranded DNA probes are prepared and labeled using [<sup>3</sup>H]TTP and terminal transferase to a suitable specific activity (e.g. approx. 420i/mmol). THAP1 protein or a portion thereof is prepared and the quantity of THAP1 protein or a portion thereof is determined via ELISA. For assay development purposes, electrophoretic mobility shift assays (EMSA) can be carried out to select suitable assay parameters. For the high throughput assay, <sup>3</sup>H labeled DNA, anti-THAP1 monoclonal antibody and THAP1 in binding buffer (Hepes, pH 7.5; EDTA; DTT; 10mM ammonium sulfate; KCl and Tween-20) are combined. The assay is configured in a standard 96-well plate and incubated at room temperature for 5 to 30 minutes, followed by the addition of 0.5 to 2 mg of PVT protein A SPA beads in 50-100 µl binding buffer. The radioactivity bound to the SPA beads is measured using a TopCount™ Microplate Counter (Packard Biosciences, Meriden, CT).

## EXAMPLE 30

High throughput in vitro assay to identify inhibitors ofTHAP-family polypeptide or THAP-family interactions with specific DNA targets

High throughput assays for the detection and quantification of THAP1 specific DNA binding is carried out using a scintillation proximity assay. Materials are available from Amersham (Piscataway, NJ) and assays can be carried out according to Gal S. et al, 6<sup>th</sup> Ann. Conf. Soc. Biomol. Screening, 6-9 Sept 2000, Vancouver, B.C.).

THAP1-specific double stranded DNA probes corresponding to THAP1 DNA binding sequences obtained according to Example 28 are prepared. The probes are labeled using [<sup>3</sup>H]TTP and terminal transferase to a suitable specific activity (e.g. approx. 420i/mmol). THAP1 protein or a portion thereof is prepared and the quantity of THAP1 protein or a portion thereof is determined via ELISA. For assay development purposes, electrophoretic mobility shift assays (EMSA) can be carried out to select suitable assay parameters. For the high throughput assay, <sup>3</sup>H labeled DNA, anti-THAP1 monoclonal antibody, 1µg non-specific DNA (double or single stranded poly-dAdT)

and THAP1 protein or a portion thereof in binding buffer (Hepes, pH7.5; EDTA; DTT; 10mM ammonium sulfate; KCl and Tween-20) are combined. The assay is configured in a standard 96-well plate and incubated at room temperature for 5 to 30 minutes, followed by the addition of 0.5 to 2 mg of PVT protein A SPA beads in 50-100µl binding buffer. The radioactivity bound to the SPA beads is measured using a TopCount™ Microplate Counter (Packard Biosciences, Meriden, CT).

### EXAMPLE 31

#### Preparation of antibody compositions

Substantially pure THAP1 protein or a portion thereof is obtained. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms per ml. Monoclonal or polyclonal antibodies to the protein can then be prepared as follows: Monoclonal Antibody Production by Hybridoma Fusion Monoclonal antibody to epitopes in the THAP1 protein or a portion thereof can be prepared from murine hybridomas according to the classical method of Kohler and Milstein (Nature, 256: 495, 1975) or derivative methods thereof (see Harlow and Lane, Antibodies A Laboratory Manual, Cold Spring Harbor Laboratory, pp. 53-242, 1988).

Briefly, a mouse is repetitively inoculated with a few micrograms of the THAP1 protein or a portion thereof over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., Meth. Enzymol. 70: 419 (1980). Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. et al. Basic Methods in Molecular Biology, Elsevier, New York., Section 21-2.

#### **Polyclonal Antibody Production by Immunization**

Polyclonal antiserum containing antibodies to heterogeneous epitopes in the THAP1 protein or a portion thereof can be prepared by immunizing suitable non-human animal with the THAP1 protein or a portion thereof, which can be unmodified or modified to enhance immunogenicity. A suitable nonhuman animal, preferably a non-human mammal, is selected. For example, the animal may be a mouse, rat, rabbit, goat, or horse. Alternatively, a crude protein preparation which, has been enriched for THAP1 or a portion thereof can be used to generate antibodies. Such proteins, fragments or preparations are introduced into the non-human mammal in the presence of an appropriate adjuvant (e. g. aluminum hydroxide, RIBI, etc.) which is known in the art. In addition the protein, fragment or preparation can be pretreated with an agent which will increase

antigenicity, such agents are known in the art and include, for example, methylated bovine serum albumin (mBSA), bovine serum albumin (BSA), Hepatitis B surface antigen, and keyhole limpet hemocyanin (KLH). Serum from the immunized animal is collected, treated and tested according to known procedures. If the serum contains polyclonal antibodies to undesired epitopes, the polyclonal antibodies can be purified by immunoaffinity chromatography.

Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987). An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al. J. Clin. Endocrinol. Metab. 33: 988-991 (1971). Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. et al., Chap. 19 in: Handbook of Experimental Immunology D. Wier (ed) Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12: M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: Manual of Clinical Immunology, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D. C. (1980).

Antibody preparations prepared according to either the monoclonal or the polyclonal protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; or they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

#### EXAMPLE 32

##### Two Hybrid THAP1/Chemokine Interaction Assay

Two-hybrid interaction between THAP1 and chemokines CCL21, CCL19, CXCL9, CXCL10, CXCL11 or cytokine IFN $\gamma$  was tested by cotransformation of AH109 with pGADT7-THAP1 and pGBKT7-CCL21, -CCL19, -CXCL9, -CXCL10, -CXCL11 and -IFN $\gamma$  plasmids and selection of transformants by His and Ade double auxotrophy according to manufacturer's instructions (MATCHMAKER two-hybrid system 3, Clontech). pGBKT7-chemokine vectors were generated using cDNAs encoding the mature forms of human chemokines CCL21 (see Example 15) (SLC polypeptide SEQ ID NO: 271, SLC cDNA SEQ ID NO: 272); CCL19 (amino acids 22-98 of GenBank Accession No. NM\_006274) (CCL19 polypeptide SEQ ID NO: 273, CCL19 cDNA SEQ ID NO: 274); CXCL9 (amino acids 23-125 of GenBank Accession No. NM\_002416) (CXCL9

polypeptide SEQ ID NO: 275, CXCL9 cDNA SEQ ID NO: 276); CXCL10 (amino acids 22-98 of GenBank Accession No. NM\_001565) (CXCL10 polypeptide SEQ ID NO: 277, CXCL10 cDNA SEQ ID NO: 278); CXCL11 (amino acids 22-94 of GenBank Accession No. NM\_005409) (CXCL11 polypeptide SEQ ID NO: 323, CXCL11 cDNA SEQ ID NO: 324) or cytokine IFN $\gamma$  (amino acids 21-166 of GenBank Accession No. NM\_000619) (IFN $\gamma$  polypeptide SEQ ID NO: 279, IFN $\gamma$  cDNA SEQ ID NO: 280), amplified by PCR, respectively, from Image clones No. 1707527 (hCCL19) with primers CCL19-1 (5'-GCGGAATCATGGGCACCAATGATGCTGAAGACTG-3') (SEQ ID NO: 281) and CCL19-2 (5'-GCGGGATCCTTAAGTCTGCTGCGGCGCTTCATCTTG-3') (SEQ ID NO: 282), No. 5228247 (hCXCL9) with primers CXCL9-1 (5'-GCCGAATTCACCCCAGTAGTGAGAAAGGGTTCGCTG-3') (SEQ ID NO: 283) and CXCL9-2 (5'-CGCGGATCCTTATGTAGTCTTCTTTTGACGAGAACGTTG-3') (SEQ ID NO: 284), No. 4274617 (hCXCL10) with primers CXCL10-1 (5'-GCCGAATTCGTACCTCTCTCTAGAACCGTACGCTGT-3') (SEQ ID NO: 285) and CXCL10-2 (5'-GCGGGATCCTTAAGGAGATCTTTTAGACATTTCCTTGCTA-3') (SEQ ID NO: 286), No. 3934139 (hCXCL11) with primers CXCL11-1 (5'-GGGGAATTCTTCCCCATGTTCAAAAGAGGAC-3') (SEQ ID NO: 325) and CXCL11-2 (5'-GGGGATCCTTAAAAATTCTTTCTTTCAAC-3') (SEQ ID NO: 326), No. 2403734 (hIFN $\gamma$ ) with primers IFN-1 (5'-GCGGAATCATGTGTTACTGCCAGGACCCATATG-3') (SEQ ID NO: 287) and IFN-2 (5'-GCGGGATCCTTACTGGGATGCTCTTCGACCTTG-3') (SEQ ID NO: 288). The PCR products were digested with EcoRI and BamHI, and cloned between EcoRI and BamHI cloning sites of vector pGBKT7 (Clontech). Positive two-hybrid interaction of THAP1 was observed with chemokines CCL21, CCL19, CXCL9 and CXCL11 while chemokine CXCL10 gave an intermediate result (+/-) in this two-hybrid assay (see Figure 19). The negative cytokine control, IFN $\gamma$ , did not have a positive interaction.

It will be appreciated that the above-described methods can be used to determine whether any particular chemokine binds to any THAP-family polypeptide. For example, cDNAs encoding THAP-family members THAP1 to THAP11 as well as THAP0 from humans and other species can be cloned into a first component vector of a two hybrid system. cDNAs encoding chemokines can be cloned into a second component vector of a two hybrid system. The two vectors can be transformed into an appropriate yeast strain, wherein the polypeptides are expressed and tested for interaction. For example, chemokine CCL5 (polypeptide SEQ ID NO: 289, cDNA SEQ ID NO: 290) can be tested for interaction with full-length THAP-1 or particular portions of THAP1, such as a nested deletion series. Chemokines which can be tested for interaction with THAP-family proteins include, but are not limited to, XCL1, XCL2, CCL1, CCL2, CCL3, CCL3L1, SCYA3L2, CCL4, CCL4L, CCL5, CCL6, CCL7, CCL8, SCYA9, SCYA10, CCL11, SCYA12, CCL13, CCL14, CCL15, CCL16, CCL17, CCL18, CCL19, CCL20, CCL21, CCL22, CCL23, CCL24,

CCL25, CCL26, CCL27, CCL28, clone 391, CARP CC-1, CCL1, CK-1, regakine-1, K203, CXCL1, CXCL1P, CXCL2, CXCL3, PF4, PF4V1, CXCL5, CXCL6, PPBP, SPBPBP, IL8, CXCL9, CXCL10, CXCL11, CXCL12, CXCL14, CXCL15, CXCL16, NAP-4, LFCA-1, Scyba, JSC, VHSV-induced protein, CX3CL1 and fCL1.

## EXAMPLE 33

*In Vitro* THAP1/Chemokine Interaction Assay

To confirm the interaction observed in yeast two-hybrid system, we performed *in vitro* GST pull down assays. THAP1, expressed as a GST-tagged fusion protein and immobilized on glutathione sepharose, was incubated with radiolabeled chemokines that were translated *in vitro*.

To generate the GST-THAP1 expression vector, the full-length coding region of THAP1 (a nucleic acid encoding amino acids 1-213 of THAP1) was amplified by PCR from HEVEC cDNA with primers *2HMR8* (5'-CGCGGATCCGTGCAGTCCTGCTCCGCCTACGGC-3') (SEQ ID NO: 291) and *2HMR11* (5'-CCGAATTCTTATGCTGGTACTTCAACTATTTCAAAGTAG-3') (SEQ ID NO: 292), digested with BamHI and EcoRI, and cloned in frame downstream of the Glutathione S-Transferase ORF, between the BamHI and EcoRI sites of the pGEX-2T prokaryotic expression vector (Amersham Pharmacia Biotech, Saclay, France). The GST-THAP1 fusion protein encoded by plasmid pGEX-2T-THAP1 and the control GST protein encoded by plasmid pGEX-2T, were then expressed in *E. Coli* DH5 $\alpha$  and purified by affinity chromatography with glutathione sepharose according to supplier's instructions (Amersham Pharmacia Biotech). The yield of proteins used in GST pull-down assays was determined by SDS-Polyacrylamide Gel Electrophoresis (PAGE) and Coomassie blue staining analysis.

*In vitro*-translated chemokines were generated with the TNT-coupled reticulocyte lysate system (Promega, Madison, WI, USA) using as templates pGBKT7-CCL21, -CCL19, -CXCL9, -CXCL10 and -CXCL11 chemokine vectors (see Example 32) or pCMV-SPORT6-CCL5 plasmid (Image clone No. 4185200). *In vitro*-translated IFN $\gamma$  cytokine was generated similarly using as template plasmid pGBKT7-IFN $\gamma$ . A 25  $\mu$ l volume of  $^{35}$ S-labelled chemokine was incubated with immobilized GST-THAP1 or GST proteins overnight at 4  $^{\circ}$ C, in the following binding buffer: 10 mM NaPO $_4$  pH 8.0, 140 mM NaCl, 3 mM MgCl $_2$ , 1 mM dithiothreitol (DTT), 0.05% NP40, and 0.2 mM phenylmethyl sulphonyl fluoride (PMSF), 1 mM Na vanadate, 50 mM  $\beta$ -glycerophosphate, 25  $\mu$ g/ml chymotrypsine, 5  $\mu$ g/ml aprotinin, and 10  $\mu$ g/ml leupeptin. Beads were then washed 5 times in 1 ml binding buffer. Bound proteins were eluted with 2X Laemmli SDS-PAGE sample buffer, fractionated by 10% SDS-PAGE and visualized by fluorography using Amplify (Amersham Pharmacia Biotech). GST/THAP1 specifically bound to chemokines CCL21, CCL19, CCL5, CXCL9, CXCL10 and CXCL11 but not cytokine IFN $\gamma$  (Figures 19 and 20). Figure 19 shows that CCL21, CCL19, CCL5, CXCL9 and CXCL11 all strongly bound to immobilized GST-THAP1 (indicated by +++ in the column labelled "*In vitro* binding to GST-THAP1"). CXCL10 also bound to immobilized GST-THAP1 (indicated by ++ in the column labelled "*In vitro* binding to GST-THAP1").

binding to GST-THAP1"). The cytokine IFN $\gamma$  did not bind to immobilized GST-THAP1 (indicated by - in the column labelled "*In vitro* binding to GST-THAP1"). Chemokines CCL21, CCL19, CCL5, CXCL9, CXCL10 and CXCL11 failed to interact with GST beads (negative control). Figure 20a shows that equivalent amounts of chemokine or cytokine were tested in the *in vitro* GST-THAP1 binding assays. Figure 20b shows that neither the cytokine, IFN $\gamma$ , nor any of the chemokines bound to immobilized GST alone. Figure 20c shows that chemokines, CXCL10, CXCL9 and CCL19, but not the cytokine IFN $\gamma$ , bound to immobilized GST-THAP1 fusions.

It will be appreciated that the above-described methods can be used to determine whether any particular chemokine binds to any THAP-family polypeptide. For example, cDNAs encoding THAP-family members THAP1 to THAP11 as well as THAP0 from humans and other species can be cloned and expressed as a GST fusion protein and immobilized to a solid support. cDNAs encoding chemokines can be translated *in vitro* and the resulting proteins incubated with the immobilized GST-THAP family fusions. Furthermore, a nested deletion series and/or point mutants of the THAP-family polypeptides can also be generated as GST-fusions and tested to determine the exact location of the chemokine binding domain for any THAP-family polypeptide with respect to any chemokine. Chemokines which can be tested for binding to THAP-family proteins include, but are not limited to, XCL1, XCL2, CCL1, CCL2, CCL3, CCL3L1, SCYA3L2, CCL4, CCL4L, CCL5, CCL6, CCL7, CCL8, SCYA9, SCYA10, CCL11, SCYA12, CCL13, CCL14, CCL15, CCL16, CCL17, CCL18, CCL19, CCL20, CCL21, CCL22, CCL23, CCL24, CCL25, CCL26, CCL27, CCL28, clone 391, CARP CC-1, CCL1, CK-1, regakine-1, K203, CXCL1, CXCL1P, CXCL2, CXCL3, PF4, PF4V1, CXCL5, CXCL6, PPBP, SPBPBP, IL8, CXCL9, CXCL10, CXCL11, CXCL12, CXCL14, CXCL15, CXCL16, NAP-4, LFCA-1, Scyba, JSC, VHSV-induced protein, CX3CL1 and fCL1.

#### EXAMPLE 34

##### Chemotaxis Bioassay: Inhibition of CCL21/CCL19-Induced

##### Chemotaxis by THAP1 Oligomeric Forms

To demonstrate inhibition of CCL21/CCL19-induced chemotaxis by THAP1 oligomers, fresh lymphocytes and a human cell line, each of which expresses the CCL21/CCL19 receptor CCR7, are assayed for a chemotactic response to chemokines in the presence or absence of oligomeric THAP1. Lymphocytes are purified from fresh heparinized human blood or mouse lymph nodes and grown in RPMI 1640 glutamax I (Invitrogen GIBCO). HuT78 cells are obtained from American Tissue Type Culture Collection (Accession Number TIB-161) and grown in Iscove's modified Dulbecco's medium with 4 mM L-Glutamine adjusted to contain 1.5g/l sodium bicarbonate (Invitrogen GIBCO). Recombinant CCL21 and CCL19 human chemokines are obtained from commercial suppliers (for example, R&D or Chemicon).

Chemokine CCL21 or CCL19 is diluted in the appropriate culture medium without serum at 20 ng/ml and 75  $\mu$ l of this solution is transferred in appropriated wells of a 96-well microplate.



Recombinant human THAP1 oligomers (GST-THAP1 or Fc-THAP1 chimera) are serially diluted starting at 500 nM and 25 µl of the different dilutions are transferred in appropriate wells. Transwells are set carefully on each well and 100 µl of a cell suspension at  $8.10^6$  cell/ml is added in the upper chamber. Following a 4-hour incubation at 37°C and 5% CO<sub>2</sub>, the cells which have migrated to the lower chamber are quantified using the Celltiter Glo system (Promega). A staining of the filter is also performed to verify that no cells adhered to the lower side of the filter after the migration. The degree of CCL21/CCL19 induced chemotaxis inhibition by THAP1 oligomeric forms is calculated by comparing the number of cells which have migrated in the presence or absence of the THAP1 oligomeric forms.

## EXAMPLE 35

Inhibition of CCL21/CCL19-Induced LymphocyteAdhesion to Endothelial Cells *In Vivo* by THAP1 Oligomeric Forms

The capacity of THAP1 oligomeric forms to block the activity of CCL21/CCL19 *in vivo*, including CCL21/CCL19-induced lymphocyte adhesion to endothelial cells, is assessed by measuring the 'rolling/sticking phenotype' of lymphocytes in mouse lymph nodes HEVs (High endothelial venules) using intravital microscopy (microscopy on live animals) as described in von Andrian (1996) Microcirculation (3):287-300; and von Andrian UH, M'Rini C. (1998) Cell Adhes Commun. 6(2-3):85-96). The rolling/sticking assay is performed as follows. In brief, the different steps of lymphocyte migration through HEVs (tethering, rolling, sticking, transendothelial migration) are analyzed by intravital microscopy in mice treated with recombinant human THAP1 oligomers (GST-THAP1 or Fc-THAP1 chimera). For observation of lymph nodes, HEVs vessels (and adhesion processes occurring in these vessels) by intravital microscopy, a microsurgical exposition of the subiliac (superficial inguinal) lymph node is made on an anaesthetized mouse. Briefly, BALB/c mice (Charles River, St Germain sur l'Arbresle, France) are anesthetized by intraperitoneal injection of 5 mg/mL ketamine and 1 mg/mL xylazine (10 mL/kg) and surgically prepared under a stereomicroscope (Leica Microsystems SA, Rueil-Malmaison, France) to allow exposure of the node vessels. A catheter is inserted in the contralateral femoral artery to permit subsequent retrograde injections of fluorescent cell suspensions or recombinant THAP1 oligomeric forms (GST-THAP1 or Fc-THAP1, 10-100 µg in 250 µl saline injected and allowed to bind for 5-30 min before injection of fluorescent cell suspensions). The mouse is then transferred to an intravital microscope (INM 100; Leica Microsystems SA). Body temperature is maintained at 37°C using a padding heater. Lymph node vessels and fluorescent cells are observed through 10 × or 20 × water immersion objective (Leica Microsystems SA) by transillumination or epifluorescence illumination. Transilluminated and fluorescent events are visualized using a silicon-intensified target camera (Hamamatsu Photonics, Massy, France) and recorded for later off-line analysis (DSR-11 Sony, IEC-ASV, Toulouse). Lymphocyte behavior in lymph node vessels is analyzed off-line as previously described (von Andrian (1996) Microcirculation (3):287-300; and von Andrian UH,

M'Rini C. (1998) Cell Adhes Commun. 6(2-3):85-96). Briefly, the rolling fraction is determined in every visible lymph node HEV as the percentage of lymphocytes interacting with the endothelial lining over the total cell number entering the venule during an observation period. Rolling cells that become subsequently adherent are included in the rolling fraction. The sticking fraction is determined as percentage of rollers that becomes firmly adherent in HEVs for more than 20 seconds. Only vessels with more than 10 rolling cells are included. The degree of inhibition of CCL21/CCL19-induced lymphocyte adhesion by THAP1 oligomers *in vivo* is calculated by comparing the number of lymphocytes sticking to endothelial cells (sticking fractions) in the presence or absence of the THAP1 oligomeric forms.

## EXAMPLE 36

Use of THAP1 Oligomeric Forms to Antagonize  
Chemokines in a Mouse Model of Rheumatoid Arthritis

This experiment is designed to test effect of antagonizing chemokines with THAP1 oligomeric forms in a mouse model of rheumatoid arthritis, the well-known collagen-induced arthritis model. In each experiment, male DBA/1 mice are immunized with collagen on day 21 and are boosted on day 0. Mice are treated daily from days 0-14 with IP injections of THAP1 oligomeric forms (GST-THAP1 or THAP1-Fc chimera) at 150, 50, 15, and 5 µg/day, and compared to mice treated with control proteins (GST or human IgG1), at 150 µg/day (n=15/group in each experiment). The incidence and severity of arthritis is monitored in a blind study. Each paw is assigned a score from 0 to 4 as follows: 0=normal; 1=swelling in 1 to 3 digits; 2=mild swelling in ankles, forepaws, or more than 3 digits; 3=moderate swelling in multiple joints; 4=severe swelling with loss of function. Each paw is totaled for a cumulative score/mouse. The cumulative scores are then totaled for mice in each group for a mean clinical score. Groups of 15 mice are treated with the indicated doses of THAP1-Fc or with 150 µg/day of human IgG1. The capacity of THAP1 oligomeric forms (GST-THAP1 or THAP1-Fc chimera) to reduce the disease incidence and severity of arthritis is determined by comparison with the control group.

## EXAMPLE 37

Use of THAP1 Oligomeric Forms to Antagonize  
Chemokines in a Mouse of Inflammatory Bowel Disease

The effect of blocking chemokines with THAP1-Fc chimera is analyzed in an experimentally induced model of Inflammatory Bowel Disease (IBD). One of the most widely used models of IBD is the DSS model (dextran sulphate salt). In this model, dextran sulphate salt (M.W. typically about 40,000 but molecular weights from 40,000 to 500,000 can be used) is given to mice (or other small mammals) in their drinking water at 2% to 8%. In some studies, C57BL/6 mice are given 2% DSS from day 0 to day 7 (D0 - D7), wherein the number of mice per group equals four (n=4). The study groups are divided as follows: No DSS + human IgG1 (250 µg/day/mouse D0 - D7); 2% DSS + THAP1-Fc (100 µg/day/mouse D0 - D7); 2% DSS + THAP1-Fc (250

µg/day/mouse D0 - D7); 2% DSS + THAP1-Fc (500 µg/day/mouse D0 - D7); 2% DSS + human IgG1 (250 µg/day/mouse D0 - D7). Mice are weighed daily. Weight loss is a clinical sign of the disease. Tissues are harvested at day 8 (D8). Histopathology is performed on the following tissues: small intestine, large intestine and mesenteric lymph nodes (MLN). The capacity of the THAP1-Fc chimera, to attenuate some of the weight loss associated with DSS-induced colitis, and to reduce inflammation in the large intestine is determined by comparing mice treated with THAP1-Fc to mice treated with control human IgG1.

#### EXAMPLE 38

##### THAP1 expression is linked to cell proliferation

To investigate the subcellular localization of endogenous THAP1, human primary endothelial cells from umbilical vein (HUVEC, PromoCell, Heidelberg, Germany) were analyzed by indirect immunofluorescence with anti-THAP1 specific antibodies. Anti-THAP1 antibodies (anti-THAP antisera) were generated in rabbits against a peptide derived from the THAP domain of human THAP1, AVRRKNFKPTKYSSIC (amino acids 39-54 in SEQ ID: 3), and affinity-purified against the corresponding peptide (Custom polyclonal antibody production, Eurogentec).

Endothelial cells were allowed to grow for 24 h to 48 h on coverslips. Cells were fixed in methanol for 5 min at -20°C, followed by incubation in cold acetone at -20°C for 30 sec. Cells were then blocked with PBS-BSA (PBS with 1% bovine serum albumin) for 10 min and then incubated 2 hr at room temperature with the rabbit polyclonal anti-THAP antibodies diluted in PBS-BSA (1/40). Cells were then washed three times 5 min at room temperature in PBS-BSA, and incubated for 1 hr with Cy3 (red fluorescence)-conjugated goat anti-rabbit IgG (1/1000, Amersham Pharmacia Biotech) secondary antibodies, diluted in PBS-BSA. Nuclei were revealed by staining with DAPI (4,6-Diamidino-2-phenylindole; 0.2 µg/ml, 10 min at room temperature). After extensive washing in PBS, samples were air dried and mounted in Mowiol. Images were collected on a Leica confocal laser scanning microscope. To verify the specificity of the immunostaining, in some experiments, the anti-THAP antibodies were pre-incubated with 2.5 µg/ml of the THAP antigenic peptide AVRRKNFKPTKYSSIC (SEQ ID NO: 293) or 2.5 µg/ml of a control peptide (QVEKLRKKLKTAQQRC (SEQ ID NO: 294).

This analysis revealed that expression of the endogenous THAP1 protein is linked to cell proliferation with very low or no expression in quiescent cells and high levels of expression in mitotic cells. Specifically, the micrographs showed that in primary human endothelial cells, expression of THAP1 was linked to the proliferation status of the cells and was preferentially observed in mitotic dividing cells. This immunostaining of mitotic cells with anti-THAP antibodies was specific since it was also observed in the presence of a control peptide but not after blocking with the THAP antigenic peptide.

#### EXAMPLE 38B

##### Cell cycle specific expression of THAP1 in S/G2-M phases

To investigate the subcellular localization of endogenous THAP1 during the cell cycle, human U2OS osteosarcoma cells (ATCC) were analyzed by indirect immunofluorescence with anti-THAP1 specific antibodies (see Example 38).

U2OS cells were allowed to grow for 24 hours on coverslips, then synchronized in different phases of the cell cycle by treatment with cell cycle inhibitors, aphidicoline (G1/S phase block, 1 µg/ml for 24h, Sigma ref A0781) or nocodazole (M phase block, 100 ng/ml for 24h, Sigma ref M1404). Cells in G1 phase were obtained 14h after release from the nocodazole block while cells in S and G2 phases were obtained 3h or 6h, respectively, after release from the aphidicolin block. Cells were fixed in methanol for 5 min at -20°C, followed by incubation in cold acetone at -20°C for 30 sec. Cells were then blocked with PBS-BSA (PBS with 1% bovine serum albumin) for 10 min and then incubated 2 hr at room temperature with the rabbit polyclonal anti-THAP antibodies diluted in PBS-BSA (1/40). Cells were then washed three times 5 min at room temperature in PBS-BSA, and incubated for 1 hr with Cy3 (red fluorescence)-conjugated goat anti-rabbit IgG (1/1000, Amersham Pharmacia Biotech) secondary antibodies, diluted in PBS-BSA. Nuclei were revealed by staining with DAPI (4,6-diamidino-2-phenylindole; 0.2 µg/ml, 10 min at room temperature). After extensive washing in PBS, samples were air dried and mounted in Mowiol. Images were collected on a Leica confocal laser scanning microscope. To verify the specificity of the immunostaining, in some experiments, the anti-THAP antibodies were pre-incubated with 2.5 µg/ml of the THAP antigenic peptide AVRRKNFKPTKYSSIC (SEQ ID NO: 293) or 2.5 µg/ml of a control peptide (QVEKLRKKLKTAQQRC (SEQ ID NO: 294).

This analysis revealed that expression of the endogenous THAP1 protein in the nucleus is cell cycle dependent. Specifically, the micrographs showed that in human U2OS osteosarcoma tumor cells, expression of THAP1 was linked to the proliferation status of the cells and was specifically observed in S/G2-M phases of the cell cycle. This immunostaining of cells in S/G2-M phases of the cell cycle with anti-THAP antibodies was specific since it was also observed in the presence of a control peptide but not after blocking with the THAP antigenic peptide.

#### EXAMPLE 39

##### THAP1 modulates transcription

To analyze the effects of THAP1 in transcriptional regulation, Gal4-luciferase reporter assays were performed. The method is carried out essentially as described in Vandel et al. (2001) Mol Cell Biol 21:6484-6494, and Vaute et al. (2002) Nucleic Acids Res 30:475-481. The full-length coding region of THAP1 (amino acids 1-213) was amplified by PCR from HEVEC cDNA with primers THAP1-Gal4.1 (5'-CCGAATTCAGGATGGTGCAGTCCTGCTCCGCCT-3') (SEQ ID NO: 295) and THAP1-Gal4.2 (5'-GCGCTCTAGATTATGCTGGTACTTCAACTATTTCAAAGTAG-3') (SEQ ID NO: 296), digested with EcoRI and XbaI, and cloned in frame downstream of the Gal4 DNA-binding domain (DBD), between the EcoRI and XbaI sites of the pCMV-Gal4 expression vector (Vandel et al.

(2001) Mol Cell Biol 21:6484-6494 ; Vaute et al. (2002) Nucleic Acids Res 30:475-481), to generate pCMV-Gal4/THAP1 expression vector. Increasing amounts of the pCMV-Gal4/THAP1 or pCMV-Gal4 expression vectors (0.025 mg, 0.05 mg, 0.1 mg, 0.2 mg, 0.5 mg, 1 mg of plasmid DNA ) were co-transfected in COS7 cells, together with a pBS-luciferase reporter plasmid (plasmid Gal4-M2-luc, 2 mg) containing four Gal4-UAS upstream of the luciferase reporter gene, and a pCMV-lacZ (0.5 mg) coding for  $\beta$ -galactosidase. A pCMV-Gal4/Suv39H1 plasmid, encoding the transcriptional repressor Suv39H1 (Vandel et al. (2001) Mol Cell Biol 21:6484-6494 ; Vaute et al. (2002) Nucleic Acids Res 30:475-481), was used as a control for transcriptional repression. 48 h after transfection, luciferase activity was measured using a luciferase reporter assay kit (Roche). Dosage of  $\beta$ -galactosidase was used to standardize transfection efficiencies.

These Gal4-luciferase reporter assays revealed that THAP1 is able to modulate transcription (Figures 21A and 21B) and exhibits transcriptional repressor properties similar to those of the transcriptional repressor Suv39H1 (Vandel et al. (2001) Mol Cell Biol 21:6484-6494 ; Vaute et al. (2002) Nucleic Acids Res 30:475-481).

#### EXAMPLE 40

##### Analysis of the subcellular localization of chemokine SLC/CCL21

To analyze the subcellular localization of the SLC/CCL21 protein, the cDNA encoding the mature form of human SLC/CCL21 (amino acids 24-134 of SEQ ID NO: 119) (GenBank Accession Number NP\_002980) is cloned in frame downstream of the Enhanced Green Fluorescent Protein (EGFP) ORF in pEGFP.C2 vector (Clontech). The pEGFP.C2-SLC/CCL21 vector is generated by subcloning the BamHI SLC/CCL21 fragment from pGBKT7-SLC/CCL21 (see example 15) into the unique BamHI cloning site of vector pEGFPC2 (Clontech). The GFP- SLC/CCL21 expression construct is then transfected into human primary endothelial cells from umbilical vein (HUVEC, PromoCell, Heidelberg, Germany). HUVEC are grown in complete ECGM medium (PromoCell, Heidelberg, Germany), plated on coverslips and transiently transfected in RPMI medium using GeneJammer transfection reagent according to manufacturer instructions (Stratagene, La Jolla, CA, USA). Analysis by fluorescence microscopy 24 hours later reveals that the GFP-SLC/CCL21 fusion protein localizes in the nucleus while GFP alone exhibits only a diffuse staining over the entire cell.

To investigate the subcellular localization of endogenous SLC/CCL21, immunohistochemistry with anti-SLC/CCL21 antibodies is performed on human tissue sections. Tissue specimens of fresh palatine tonsils are embedded in OCT compound (TissueTek, Elkhart, IN) and then snap-frozen in liquid nitrogen. Cryosections (6  $\mu$ m) are air-dried overnight, and acetone fixed (10 min, -20°C). Following one PBS wash, sections are treated 5 min at room temperature in PBS containing 0.1% Triton-X100, and washed again with PBS. The tissue sections are then incubated with a mixture of rabbit polyclonal antibodies against human SLC/CCL21 (1/100, Chemicon, USA) followed by a mixture of Cy3-conjugated goat anti-rabbit IgG (1/1000, Amersham Pharmacia Biotech) secondary antibodies, diluted in PBS-BSA. Nuclei are revealed by

staining with DAPI (4,6-diamidino-2-phenylindole; 0.2 µg/ml, 10 min at room temperature). After extensive washing in PBS, samples are air dried and mounted in Mowiol. Microscopy is performed with a Nikon Eclipse TE300 fluorescence microscope equipped with a Nikon digital camera DXM1200 (Nikon Corp., Tokyo, Japan).

5 Experiments similar to those described above were performed in HeLa cells and GFP-SLC was shown to localize to the nucleus. Figure 27A shows areas of localization of GFP-SLC which correspond to nuclei as shown by DAPI counterstain (Figure 27B).

#### EXAMPLE 40B

##### Analysis of the subcellular localization of chemokine MIG/CXCL9

10 To analyze the subcellular localization of the MIG/CXCL9 protein, the cDNA encoding the mature form of human MIG/CXCL9 (amino acids 23-125 of GenBank Accession No. NM\_002416) (CXCL9 polypeptide SEQ ID NO: 275, CXCL9 cDNA SEQ ID NO: 276) is cloned in frame downstream of the Enhanced Green Fluorescent Protein (EGFP) ORF in pEGFP.C2 vector (Clontech). The pEGFP.C2-MIG/CXCL9 vector is generated by subcloning the EcoRI-BamHI  
15 MIG/CXCL9 fragment from pGBKT7-MIG/CXCL9 (see example 32) between the EcoRI-BamHI cloning sites of vector pEGFPC2 (Clontech). The GFP- MIG/CXCL9 expression construct is then transfected into human primary endothelial cells from umbilical vein (HUVEC, PromoCell, Heidelberg, Germany) or human immortalized Hela cells. HUVEC are grown in complete ECGM medium (PromoCell, Heidelberg, Germany), plated on coverslips and transiently transfected in  
20 RPMI medium using GeneJammer transfection reagent according to manufacturer instructions (Stratagene, La Jolla, CA, USA). Human Hela cells (ATCC) were grown in Dulbecco's Modified Eagle's Medium supplemented with 10% Fetal Calf Serum and 1% Penicillin-streptomycin (all from Life Technologies, Grand Island, NY, USA), plated on coverslips, and transiently transfected with calcium phosphate method using 2 µg pEGFPC2-MIG/CXCL9 plasmid. Analysis of  
25 transfected HUVEC or Hela cells by fluorescence microscopy 24 hours later revealed that the GFP-MIG/CXCL9 fusion protein, similarly to GFP-SLC/CCL21 localizes in the nucleus while GFP alone exhibits only a diffuse staining over the entire cell.

Experiments similar to those described above were performed in HeLa cells and GFP-MIG was shown to localize to the nucleus. Figure 27C shows areas of localization of GFP-MIG which  
30 correspond to nuclei as shown by DAPI counterstain (Figure 27D).

#### EXAMPLE 40C

##### CXCR3-dependent nuclear translocation of chemokine MIG/CXCL9

To demonstrate the capacity of secreted chemokine MIG/CXCL9 to undergo CXCR3-dependent nuclear translocation, the cDNA encoding the full length form of human MIG/CXCL9 (amino acids  
35 1-125 of GenBank Accession No. NM\_002416) (CXCL9 polypeptide SEQ ID NO: 275, CXCL9 cDNA SEQ ID NO: 276) was amplified by PCR from Image clone No. 5228247 with primers CXCL9-3 (5'-CCGAATCCCCACCATGAAGAAAAGTGGTGTCTTT-3') (SEQ ID NO: 327)

and CXCL9-4 (5'-CCGGATCCTGTAGTCTTCTTTTGACGAGAACGTTG-3') (SEQ ID NO: 328), digested with EcoRI and BamHI, and cloned between EcoRI and BamHI cloning sites of vector pFLAG-CMV-5a (Sigma) to generate the pHMIG-Flag expression vector. The CXCR3 expression vector pEF-CXCR3 was generated by cloning the full length CXCR3 cDNA (amino acids 1-368 of GenBank Accession No. NM\_001504) (CXCR3 polypeptide SEQ ID NO: 304, CXCR3 cDNA SEQ ID NO: 305), amplified by PCR from Image clone No. 5176136 with primers 5'Xba-CXCR3 (5'-CCTCTAGACCACCATGGTCCTTGAGGTGAGTGAC-3') (SEQ ID NO: 329) and 3'Not-CXCR3 (5'-CCCGCGGCCGCTCACAAGCCCGAGTAGGAGGC-3') (SEQ ID NO: 330), between the XbaI and NotI sites of the pEF-BOS expression vector (Mizushima and Nagata, Nucleic Acids Research, 18:5322, 1990). The pHMIG-Flag expression construct was then transfected into human U2OS osteosarcoma cancer cells. Human U2OS cells (ATCC) were grown in Dulbecco's Modified Eagle's Medium supplemented with 10% Fetal Calf Serum and 1% Penicillin-streptomycin (all from Life Technologies, Grand Island, NY, USA), plated on coverslips, and transiently transfected with calcium phosphate method using 2 µg pHMIG-Flag plasmid together with pEF-CXCR3 or pEF-Bos control vector. U2OS cells co-transfected with pHMIG-Flag and pEF-CXCR3 or pEF-Bos expression vectors were analysed 48h later by fluorescence microscopy. Cells were washed twice with PBS, fixed for 15 min at room temperature in PBS containing 3.7% formaldehyde, and washed again with PBS prior to neutralization with 50mM NH<sub>4</sub>Cl in PBS for 5 min at room temperature. Following one more PBS wash, cells were permeabilized 5 min at room temperature in PBS containing 0.1% Triton-X100, and washed again with PBS. Permeabilized cells were then blocked with PBS-BSA (PBS with 1% bovine serum albumin) for 10 min and then incubated 2 hr at room temperature with rabbit polyclonal antibodies anti-Flag epitope (1/200, Sigma) and mouse monoclonal antibody anti-CXCR3 (mouse IgG1, 1/200, R&D) diluted in PBS-BSA. Cells were then washed three times 5 min at room temperature in PBS-BSA, and incubated for 1 hr with Cy3 (red fluorescence)-conjugated goat anti-rabbit IgG (1/1000, Amersham Pharmacia Biotech) and FITC-labeled goat anti-mouse-IgG (1/40, Zymed Laboratories Inc., San Francisco, CA, USA) secondary antibodies, diluted in PBS-BSA. After extensive washing in PBS, samples were air dried and mounted in Mowiol. Images were collected on a Leica confocal laser scanning microscope. The FITC (green) and Cy3 (red) fluorescence signals were recorded sequentially for identical image fields to avoid cross-talk between the channels.

In cells co-transfected with pHMIG-Flag and pEF-CXCR3 expression vectors, hMIG-Flag was found to accumulate in the nucleus of the majority of transfected cells (Figures 28A-D and 29A-C). Nuclear localization of MIG-Flag was specifically observed in CXCR3-positive cells (Figure 29A-C) and was not found in cells co-transfected with the pEF-Bos control vector (Figure 28A-D). These results demonstrated that chemokine receptor CXCR3 plays an essential role in nuclear translocation of secreted chemokine MIG.

## EXAMPLE 41

The THAP1/SLC-CCL21 complex modulates transcription

To analyze the effects of SLC/CCL21 and the THAP1/SLC-CCL21 complex in transcriptional regulation, Gal4-luciferase reporter assays are performed essentially as described in Example 39. The SLC/CCL21 expression vector used in these transcription assays (pEF-SLC/CCL21) is generated by PCR. A cDNA encoding the mature form of human SLC/CCL21 (amino acids 24-134 of SEQ ID NO: 119) (GenBank Accession Number NP\_002980), is amplified by PCR from HEVEC RNA with primers hSLC.Xba (5'-GCGTCTAGAATGAGTGATGGAGGGGCTCAGGACTGTTG-3') (SEQ ID NO: 297) and hSLC.Not (5'-GGGCGGCCGCCTATGGCCCTTTAGGGGTCTGTGACCGC-3') (SEQ ID NO: 298), digested with XbaI and NotI, and cloned into the XbaI and NotI sites of the pEF-BOS expression vector (Mizushima and Nagata, Nucleic Acids Research, 18:5322, 1990).

Increasing amounts of the pEF-SLC/CCL21 plasmid (0.025 mg, 0.05 mg, 0.1 mg, 0.2 mg, 0.5 mg, 1 mg of plasmid DNA) are co-transfected in COS7 cells, together with pCMV-Gal4/THAP1 or pCMV-Gal4 expression vectors (0.5 mg), a pBS-luciferase reporter plasmid (plasmid Gal4-M2-luc, 2 mg) containing four Gal4-UAS upstream of the luciferase reporter gene, and a pCMV-lacZ (0.5 mg) coding for  $\beta$ -galactosidase. Forty-eight hours after transfection, luciferase activity is measured using a luciferase reporter assay kit (Roche). Dosage of  $\beta$ -galactosidase is used to standardize transfection efficiencies.

These Gal4-luciferase reporter assays should reveal that SLC/CCL21 is able to modulate transcriptional activity of THAP1, indicating a role for the THAP1/SLC-CCL21 complex in transcriptional regulation (Figure 22A).

Similar to other cytokines such as IFN $\gamma$  (Bader and Wietzerbin (1994) PNAS 91:11831-11835; Subramaniam et al. (1999) J Biol Chem 274:403-407) and growth factors such as FGF2 (Baldin et al. (1990) EMBO J 9:1511-1517), the basic SLC/CCL21 chemokine may be internalized and translocated to the nucleus, where it may associate with THAP1 and modulate (stimulate or inhibit) transcription of specific target genes. Target genes of THAP1 and THAP1/SLC complex can include genes involved in cell proliferation and cell differentiation, particularly endothelial cell differentiation and endothelial or cancer cell proliferation.

It will be appreciated that the above-described methods can be used to determine whether any particular THAP1/chemokine complex or THAP-family polypeptide/chemokine complex has the ability to modulate transcription. For example, cDNAs encoding THAP-family members THAP1 to THAP11 as well as THAP0 from humans and other species can be cloned in an expression vector such as pCMV-Gal4, the desired chemokine is cloned into the expression vector pEF-BOS and the expression constructs are then either transfected separately or cotransfected into COS7 cells comprising a pBS-luciferase reporter plasmid. Luciferase assays are performed as described above.



Chemokines which can be tested in combination with THAP1 or other THAP-family polypeptides for their ability to modulate transcription include, but are not limited to, XCL1, XCL2, CCL1, CCL2, CCL3, CCL3L1, SCYA3L2, CCL4, CCL4L, CCL5, CCL6, CCL7, CCL8, SCYA9, SCYA10, CCL11, SCYA12, CCL13, CCL14, CCL15, CCL16, CCL17, CCL18, CCL19, CCL20, CCL21, CCL22, CCL23, CCL24, CCL25, CCL26, CCL27, CCL28, clone 391, CARP CC-1, CCL1, CK-1, regakine-1, K203, CXCL1, CXCL1P, CXCL2, CXCL3, PF4, PF4V1, CXCL5, CXCL6, PPBP, SPBPBP, IL8, CXCL9, CXCL10, CXCL11, CXCL12, CXCL14, CXCL15, CXCL16, NAP-4, LFCA-1, Scyba, JSC, VHSV-induced protein, CX3CL1 and fCL1.

In experiments conducted using MIG and THAP1, it was shown that MIG/THAP1 complexes could modulate gene transcription (see Figure 22B and Example 47).

#### EXAMPLE 42

##### *Fucosyltransferase TVII* is a target gene of THAP1 and/or the THAP1/SLC-CCL21 complex

Since chemokine SLC/CCL21 has been shown to induce the high endothelial venule phenotype in endothelial cells (Fan et al. (2000) J Immunol 164:3955-3959; Grant et al. (2002) Am J Pathol 2002 160:1445-55; Yoneyama et al. (2001) J Exp Med 193:35-49), we searched for target genes of the THAP1/SLC-CCL21 among the few high endothelial venule-specific genes that have been described. This analysis led to the identification of many THAP domain recognition sequences in the promoter of the human *Fucosyltransferase TVII* gene (Figure 23), one of the key high endothelial venules enzymes for lymphocyte recruitment (Smith et al. (1996) J Biol Chem 271:8250-8259; Maly et al. (1996) Cell 86:643-653).

To confirm that the *Fucosyltransferase TVII* promoter is a target of THAP1 and/or the THAP1/SLC-CCL21 complex, transcription assays are performed with luciferase reporter genes under the control of the *FucTVII* promoter. The *FucTVII* promoter (nucleotides 650-950, GenBank Accession Number AB012668) is amplified by PCR from human genomic DNA with primers *FucTVII*-1 (5'-GCGCTCGAGCTGCACCTGGGCCTTCTCTGCCCTGG-3') (SEQ ID NO: 299) and *FucTVII*-2 (5'-CGAAGCTTACTGTGCTCCTTTTATCTCTGCCCAAG-3') (SEQ ID NO: 300), digested with XhoI and HindIII, and cloned in the same sites of the pGL3-Basic luciferase reporter plasmid (Promega) to generate pGL3-*proFucTVII*-luc.

To analyze the effects of SLC/CCL21 and the THAP1/SLC-CCL21 complex on the *FucTVII* promoter, luciferase reporter assays are performed essentially as described in Example 39. Increasing amounts of the pEF-SLC/CCL21 and/or pEGFPC2-THAP1 plasmid (0.025 mg, 0.05 mg, 0.1 mg, 0.2 mg, 0.5 mg, 1 mg of plasmid DNA) are co-transfected in COS7 cells, together with the pGL3-*proFucTVII*-luciferase reporter plasmid, and pCMV-lacZ (0.5 mg) coding for  $\beta$ -galactosidase. Forty-eight hours after transfection, luciferase activity is measured using a luciferase reporter assay kit (Roche). Dosage of  $\beta$ -galactosidase is used to standardize transfection efficiencies.

These luciferase reporter assays with the pGL3-*proFucTVII*-luciferase reporter plasmid reveals that THAP1, SLC/CCL21 and the THAP1/SLC-CCL21 complex are able to modulate transcriptional activity of the *FucTVII* promoter, indicating that the human *Fucosyltransferase TVII* gene is a target of THAP1 and the THAP1/SLC-CCL21 complex, further confirming the role of the THAP1/SLC-CCL21 complex in transcriptional regulation.

## EXAMPLE 43

Retrovirus mediated expression of THAP1 andchemokines SLC/CCL21 and MIG/CXCL9 in primary human endothelial cells

*Background:* The method described below uses retroviral derived vectors to transduce at high efficiency human primary umbilical vein endothelial cells (HUVEC) with THAP1, SLC/CCL21, MIG/CXCL9 or any other gene of interest. This retroviral packaging system includes retroviral packaging plasmids and packagable vector transcripts that are produced from high expression plasmids after transient tri-transfection in mammalian cells. High titers of recombinant retroviruses are produced in these transfected mammalian cells and can then transduce a mammalian target cell, so-called HUVEC, by fresh supernatant infection at high efficiency. This method is useful for the rapid production of high titer viral supernatants to transduce with high efficiency cells that are refractory to transduction by conventional means such as simple transfection. The transduction protocol in primary HUVEC has been optimized with a MLV-derived vector encoding enhanced green fluorescent protein (eGFP) to determine the optimal infection conditions.

The retroviral constructs are packaging plasmids consisting of at least one retroviral helper DNA sequence derived from a replication-incompetent retroviral genome encoding in trans all virion proteins required to package a replication incompetent retroviral vector, and for producing virion proteins capable of packaging the replication-incompetent retroviral vector at high titer, without the production of replication-competent helper virus. The retroviral DNA sequence lacks the region encoding the native enhancer and/or promoter of the viral 5' LTR of the virus, and lacks both the psi function sequence responsible for packaging helper genome and the 3' LTR, but encodes the foreign  $\beta$ -globin polyadenylation site. The retrovirus is a leukemia virus, the Moloney Murine Leukemia Virus (MMLV). The foreign enhancer and promoter is the human cytomegalovirus (HCMV) immediate early (IE) enhancer and promoter. The retroviral packaging plasmid consists of two retroviral helper DNA sequences encoded by plasmid based expression vectors, for example where a first helper sequence contains a cDNA encoding the gag and pol proteins of ecotropic MMLV and a second helper sequence contains a cDNA encoding the env protein. The Env gene, which determines the host range, is derived from the Vesicular Stomatitis Virus (VSV) G protein.

*Plasmid constructions:* MLV retroviral vectors were based on MoMLV derived vector called pCFB from Stratagene where the U3 region of the 5'LTR were replaced by the

enhancer/promoter of the cytomegalovirus immediate early (CMV IE) gene. The multi-cloning site was modified by incorporation of synthetic oligonucleotides 5'-GGCATTCAATTGCTCGAGTTTAAACGCGGCCGCG-3' (SEQ ID NO: 331) and 5'-AATCCGCGGCCGCGTTTAAACTCGAGCAATTGAATGCC-3' (SEQ ID NO: 332) containing the NaeI and MfeI restriction sites and replacing nucleotides from position 1742 to 2244 of the pCFB plasmid. The modified vector was called pMLV-MCS. The pVSVG plasmid encoding the VSVG envelope and the pGAGPOL plasmid encoding gag and pol genes have been constructed as follows: VSVG and GAG-POL DNA fragments were amplified from respectively pVPack-VSV-G and pVPack-GP plasmids as templates (Stratagene) and cloned into the CMV- $\beta$ globin intron-MCS- $\beta$ globin polyA expression cassette following conventional cloning procedures. Primers used to amplify vsvg and gagpol fragments were respectively VSVG-5' (5'-ATGAAGTGCCTTTTGTACTTAGCCTT-3') (SEQ ID NO: 333) and VSVG-3' (5'-TCATAAAAATTAATAAACTCAAATATAATTGAGG-3') (SEQ ID NO: 334) and GAGPOL-5' (5'-ATGGGCCAGACTGTTACCACTC-3') (SEQ ID NO: 335) and GAGPOL-3' (5'-TTAGGGGGCCTCGCGG-3') (SEQ ID NO: 336).

The full length coding region of human THAP1 (SEQ ID NO: 3; amino acids 1 to 213), were amplified by PCR according to standard procedures with primers: THAP1-5' (5'-ATGGTGCAGTCCTGCTCCGC-3') (SEQ ID NO: 337) and THAP1-MfeI-3' (5'-GCCAATTGTTATGCTGGTACTTCAACTATTT-3') (SEQ ID NO: 338) using a recombinant vector containing the human THAP1 cDNA as template. The reverse primer contains an MfeI restriction site at its end to generate a 3' overhang compatible with the 5' end of the cleaved vector. The amplified DNA were then digested with MfeI, purified by electrophoresis on an agarose gel and the desired fragments were then cloned into the cleaved vector pMLV-MCS digested with NaeI and MfeI restriction enzymes.

Coding regions of human SLC/CCL21 (Genbank NP) and human MIG/CXCL9 (NM\_002416) were amplified by PCR in such a way that the amplified fragments did not contain the signal peptide localized from the nucleotide 4 to the nucleotide 66 of the full length open reading frame of both sequences. By deleting signal peptide signatures, SLC/CCL21 and MIG/CXCL9 proteins are localized into the nucleus of the cell after protein expression in the cytoplasm. Primers used were SLC-5' (5'-ATGAGTGATGGAGGGGCTCAGG-3') (SEQ ID NO: 339) and SLC-EcoRI-3' (5'-GGAATTCCTATGGCCCTTTAGGG-3') (SEQ ID NO: 340), MIG-5' (5'-ATGACCCAGTAGTGAGAAAGGGTC-3') (SEQ ID NO: 341) and MIG-EcoRI-3' (5'-GGAATTCCTTATGTAGTCTTCTTTTGACGAGA-3') (SEQ ID NO: 342) for SLC/CCL21 and MIG/CXCL9, respectively. Both reverse primers contain an EcoRI restriction site at their end to generate a 3' overhang compatible with the 5' end of the cleaved vector. The amplified DNAs were then digested with EcoRI, purified by electrophoresis on an agarose gel and the desired fragments were then cloned into the cleaved vector pMLV-MCS digested with NaeI and EcoRI restriction

enzymes. The recombinant vectors thus obtained, pMLV-THAP1, pMLV-SLC/CCL21, pMLV-MIG/CXCL9, encode amino acids-2 to 213 of the THAP1 sequence or amino acids-24 to 134 of the matured SLC/CCL21 sequence or amino acids-23 to 125 of the matured MIG/CXCL9 sequence.

5        *Transfection, virus harvest, and retroviral infection of cells:* Retroviral vectors carrying either THAP1 or SLC/CCL21 or MIG and driven by the moloney murine leukemia virus LTR were produced by transient transfection in 293T cells (ATCC No. CRL11268, ATCC, Rockville, Md) with the following plasmids: the packaging plasmid (pGAGPOL), the envelope plasmid coding for the vesicular stomatitis virus G protein (pVSV-G), and one of the transducing vector  
10        pMLV-THAP1, pMLV-SLC, pMLV-MIG, pMLV-MCS or pMLV-EGFP. 293T cells were transfected via the calcium phosphate precipitation method (Pear et al., 1993). Briefly, cells were plated at a density of  $4 \times 10^6$  cells per 75-  $\text{cm}^2$  dishes one day prior to transfection. DNA-calcium phosphate complexes consisting of pVSVg, pGAGPOL and one of the transducing vector pMLV-THAP1, pMLV-SLC, pMLV-MIG, pMLV-MCS or pMLV-EGFP were diluted in an equal volume  
15        of HBS2x buffer and incubated with cells for 16 hours. After media removal, cells were replenished with fresh medium and further incubated for 24 hours. Cell supernatants containing viral particles were harvested every 8-12 hours, clarified of cell debris using low-speed centrifugation and filtered on 0.45  $\mu\text{m}$  filters.

*Cell Transduction:* A total of  $10^6$  HUVEC were transduced in a 75  $\text{cm}^2$  plate with 10 ml  
20        of viral supernatant in the presence of 8  $\mu\text{g}/\text{ml}$  of polybren (Sigma) as previously described (Yu. H. et al., Gene Therapy, 6, 1876-1883, 1999). After 4 hours, the supernatant was replaced by fresh endothelial cell medium consisting of MCDB131 (Gibco Brl) supplemented with 20% heat-inactivated serum, endothelial cell growth factor (ECGS, Sigma Chemical Co.) and 5 U/ml sodium heparin. When applicable, second transduction were processed using the same protocol a day after  
25        the first transduction. Forty-eight hours after the second infection, cells were trypsinized and pelleted for RNA preparation. Total RNA was isolated from  $10^6$  cells with the Absolutely RNA miniprep kit according to the manufacturer's instructions (Stratagene, La Jolla, CA, USA).

#### EXAMPLE 44

##### Identification of THAP1 target genes by DNA microarrays 30        and real-time Polymerase Chain Reaction (PCR)

      To better understand the function of THAP1 as a nuclear factor in vasculature, we globally profiled THAP1 target genes either in primary human endothelial cells or in primary endothelial cells constitutively expressing chemokines in the nucleus using retroviral gene transfer and Agilent oligonucleotide-based microarray technology. We quantitated the THAP1 mediated changes in  
35        expression of more than 17,000 mRNAs by transducing human vascular endothelial cells with the following set of viral particles: MCS as the negative infection control, THAP1, SLC/CCL21 and

enzymes. The recombinant vectors thus obtained, pMLV-THAP1, pMLV-SLC/CCL21, pMLV-MIG/CXCL9, encode amino acids-2 to 213 of the THAP1 sequence or amino acids-24 to 134 of the matured SLC/CCL21 sequence or amino acids-23 to 125 of the matured MIG/CXCL9 sequence.

5        *Transfection, virus harvest, and retroviral infection of cells:* Retroviral vectors carrying either THAP1 or SLC/CCL21 or MIG and driven by the moloney murine leukemia virus LTR were produced by transient transfection in 293T cells (ATCC No. CRL11268, ATCC, Rockville, Md) with the following plasmids: the packaging plasmid (pGAGPOL), the envelope plasmid coding for the vesicular stomatitis virus G protein (pVSV-G), and one of the transducing vector  
10    pMLV-THAP1, pMLV-SLC, pMLV-MIG, pMLV-MCS or pMLV-EGFP. 293T cells were transfected via the calcium phosphate precipitation method (Pear et al., 1993). Briefly, cells were plated at a density of  $4 \times 10^6$  cells per 75- cm<sup>2</sup> dishes one day prior to transfection. DNA-calcium phosphate complexes consisting of pVSVg, pGAGPOL and one of the transducing vector pMLV-THAP1, pMLV-SLC, pMLV-MIG, pMLV-MCS or pMLV-EGFP were diluted in an equal volume  
15    of HBS2x buffer and incubated with cells for 16 hours. After media removal, cells were replenished with fresh medium and further incubated for 24 hours. Cell supernatants containing viral particles were harvested every 8-12 hours, clarified of cell debris using low-speed centrifugation and filtered on 0.45 µm filters.

*Cell Transduction:* A total of  $10^6$  HUVEC were transduced in a 75 cm<sup>2</sup> plate with 10 ml  
20    of viral supernatant in the presence of 8 µg/ml of polybren (Sigma) as previously described (Yu. H. et al., Gene Therapy, 6, 1876-1883, 1999). After 4 hours, the supernatant was replaced by fresh endothelial cell medium consisting of MCDB131 (Gibco Brl) supplemented with 20% heat-inactivated serum, endothelial cell growth factor (ECGS, Sigma Chemical Co.) and 5 U/ml sodium heparin. When applicable, second transduction were processed using the same protocol a day after  
25    the first transduction. Forty-eight hours after the second infection, cells were trypsinized and pelleted for RNA preparation. Total RNA was isolated from  $10^6$  cells with the Absolutely RNA miniprep kit according to the manufacturer's instructions (Stratagene, La Jolla, CA, USA).

#### EXAMPLE 44

##### Identification of THAP1 target genes by DNA microarrays 30        and real-time Polymerase Chain Reaction (PCR)

      To better understand the function of THAP1 as a nuclear factor in vasculature, we globally profiled THAP1 target genes either in primary human endothelial cells or in primary endothelial cells constitutively expressing chemokines in the nucleus using retroviral gene transfer and Agilent oligonucleotide-based microarray technology. We quantitated the THAP1 mediated changes in  
35    expression of more than 17,000 mRNAs by transducing human vascular endothelial cells with the following set of viral particles: MCS as the negative infection control, THAP1, SLC/CCL21 and

MIG/CXCL9. In addition, SLC/CCL21 and MIG/CXCL9 infected endothelial cells were re-infected a day after with viral particles containing either MCS or THAP1. After 50 and 120 hours of the second infection, HUVECs cells were pelleted, washed and lysed to prepare total RNA and protein extracts. Over-expression of THAP1, SLC/CCL21 and MIG/CXCL9 in HUVECs was  
5 verified both at RNA and protein levels with standard quantitative PCR and Western blotting procedures.

*Oligonucleotide Array Expression Analysis*

Total RNA quality control was performed by running 25-50 ng on an RNA 6000 Nano Assay (Agilent) using a Bio-analyser 2100. For labelling, 500 ng of good quality total RNA was  
10 reverse-transcribed with an oligo-dT-T7 and double stranded cDNA was generated with the superscript double stranded cDNA synthesis kit (Invitrogen). In an *in vitro* transcription step with T7 RNA polymerase, the cDNA was linearly amplified and labeled with fluorescent nucleotides (low RNA input fluorescent linear amplification kit, Agilent). Ten micrograms of labeled and fragmented cRNA was then hybridized onto a Human genome 1A expression array (G4110A,  
15 Agilent) for 16 hours at 45°C. Post-hybridization staining and washing were performed according to manufacturer instructions. Finally, chips were scanned with an Agilent DNA microarray scanner at the Microarray Facility. Data acquisition and analysis were performed with the Agilent Feature Extraction and Analysis software using the Rosetta Resolver data analysis system.

*Real-Time Polymerase Chain Reaction (PCR)*

Real-Time PCR was performed on cDNA synthesized from RNA isolated from HUVEC cells infected with THAP1, MCS, SLC, MIG, SLC and MCS, SLC and THAP1, MIG and MCS or  
20 MIG and THAP1 retroviral constructs using the ABI7700 Prism SDS Real-Time PCR Detection System (Applied Biosystems). The ABI7700 Prism was formatted for 96 well plates containing 25 µl PCR reactions. Real-time PCR were made such as each 25 µl contained 2 µl DNA-template  
25 (dilution 1:4), 12.5 µl SYBR Green PCR Master Mix kit (Applied Biosystems, Foster City, CA, USA) and 0.8 µM forward and reverse gene primers. PCR conditions consisted of an initial denaturing step for 10 minutes at 95°C, followed by 40 cycles of a 2-segment step consisting of denaturation for 30 seconds at 95°C, annealing and elongation for 60 seconds at 60°C. After real  
time analysis, a melting curve was established for all samples to insure specific amplification. A  
30 negative control, where no template DNA was used, was run on each plate as well as a comparison of GAPDH between all samples. GAPDH served to equilibrate the starting material between the two experimental conditions. All unknown samples as well as controls were run in duplicate on the same plate (except for the negative control). Reactions were recorded and analysed with the  
ABI7700 Prism SDS sequence detection system. The threshold cycle ( $C_t$ ) for each sample run in  
35 duplicate were determined and fold differences was performed as detailed previously (Van Trappen et al., 2001).

## EXAMPLE 45

THAP1 regulates cell cycle specific genes and modulates  
proliferation of both primary endothelial cells and immortalized cancer cells

5 We combined data from independent microarray analyses of the effects of THAP1 on gene  
expression in primary human endothelial cells to identify THAP1-target genes (THAP responsive  
genes).

10 Table 2A lists certain genes regulated by THAP1, as revealed by two independent  
microarray experiments with human primary endothelial cells transduced with THAP1 (THAP) or  
control (MCS) retrovirus expression vectors. The GenBank and Swiss Prot accession numbers are  
indicated for each gene as well as the fold changes, p-values and signal intensities obtained in the  
two microarray experiments.

Table 2B lists database accession numbers for each gene and corresponding polypeptide  
listed in Table 2A. Table 2B also shows the Aligent oligonucleotide which corresponds to each  
gene listed in Table 2A.

15

TABLE 2A

Sequence Name(s)	Sequence Description	QUERY - Fold Change	QUERY - P-value	QUERY - Intensity1	QUERY - Intensity2	TARGET - Fold Change	TARGET - P-value	TARGET - Intensity1	TARGET - Intensity2
Name	Description	SwissProt Acc Number	Fold change 1st expt THAP vs MCS 50h	Intensity1	Intensity2	Fold change 2nd expt THAP vs MCS 50h	pvalue	Intensity1	Intensity2
USP16	Ubiquitin specific protease 16, deubiquitinates histone H2A and H2B, may play a role in deubiquitination of proteins involved in the condensation of mitotic chromosomes and may deubiquitinate histone H2A during apoptosis	Q9Y5T5	-3.37	2445.9	725.44	-1.66	6.49E-10	1224.1	739.65
CDCA7	Protein of unknown function	AAH27966.1	-2.55	1623.4	649.45	-1.6	1.53E-07	2114.2	1321.11
CKS1	CDC28 protein kinase 1, binds and regulates CDK2-cyclin A complexes, similar to S. pombe p13suc1, required for SCF dependent degradation of p27	P33551	-2.19	28194	12832	-2.01	2.20E-24	27653	13620.69
THAP1	THAP domain protein 1	BAA91635.1	-2.13	1018	471.4	-1.92	5.96E-17	1510.6	791.48
MAD2L1	Mitotic arrest deficient 2 yeast homolog-like 1, essential for timing of anaphase onset, involved in detection of kinetochore attachment as part of the mitotic spindle checkpoint; genetic variants may be linked to breast cancer	Q13257	-1.91	3934.9	2055.1	-1.69	1.36E-05	3694.8	2191.56
KIAA0008	Member of the guanylate-kinase-associated protein (GKAP) protein family	Q15398	-1.9	6136.3	3232.8	-1.54	1.36E-24	5911.6	3848.34
PTTG1	Pituitary tumor-transforming 1 (securin), a transcriptional activator that promotes cell proliferation and angiogenesis, involved in sister chromatid separation and euploidy maintenance; overexpression promotes cellular transformation and tumorigenesis	CAA11683.1	-1.63	27920	17156	-1.53	3.65E-17	20405	13308.98
BIRC5	Survivin, a member of the inhibitor of apoptosis protein family that is involved in	O15392	-1.62	24865	15286	-1.55	2.90E-08	17927	11605.75



Sequence Name(s)	Sequence Description	QUERY_Fold Change	QUERY_P-value	QUERY_Intensity1	QUERY_Intensity2	TARGET_Fold Change	TARGET_P-value	TARGET_Intensity1	TARGET_Intensity2
	G2-M transition and exit of the mitotic cell cycle; may play a role in oncogenesis								
PTTG3	Pituitary tumor-transforming 3, a protein that may be associated with tumorigenesis	-1.62	4.56E-05	26737	16489	-1.54	2.41E-24	20058	13058.54
HMMR	Hyaluronan mediated motility receptor, binds hyaluronan and is important for cell motility, binds microtubules and microfilaments intracellularly, may also be involved in cell proliferation; gene mutations may contribute to colon cancer development	-1.6	1.08E-05	3584.9	2231.3	-1.69	9.91E-06	2255.3	1286.07
PTTG2	Pituitary tumor transforming gene 2, protein with strong similarity to human PTTG1, which is a proto-oncoprotein with PXXP motifs that may bind to SH3 domains, suggesting roles in intracellular signal transduction and tumor-specific pathogenesis	-1.52	4.44E-06	4283.3	2653.2	-1.56	8.44E-07	4727.5	3064.92

TABLE 2B

TABLE 2B - Part I

Gene Name	Agilent	Protein	LocusId Symbol	RefSeq	Genomic hit	Unigene	GenBank	Alias
USP16	= I_962079	SP:Q9Y5T5	10600 USP16	NM_006447	NT_011512	Hs.99819	AF126736 m; AK023247 m; AK025104 m; none p	Ubp-M
CDCA7	= I_928296	GP:AAH27966.1	83879 CDCA7	NM_031942; NM_145810	NT_005403	Hs.333893	AK027628 m; AK027642 m; AK075134 m; AL833728 m; AL834186 m; AY029179 m; BC015124 m; BC027966 m	JPO1; FLJ14722; FLJ14736; MGC34109
CKS1	= I_929087	SP:P33551	1163 CKS1B	NM_001826	NT_004668	Hs.348669	AF279897 m; BC007751 m; BC015629 m; X54941 m; none p	CKS1; ckshs1
THAP1	= I_929644	GP:BAA91635.1	55145 THAP1	NM_018105	NT_008251	Hs.7432	AK001339 m; BC021721 m	THAP1
MAD2L1	= I_957747	SP:Q13257	4085 MAD2L1	NM_002358	NT_016354	Hs.79078	AJ000186 m; BC000356 m; BC005945 m; U31278 m; none p	MAD2; HSMAD2
KIAA0008	= I_959284	SP:Q15398	9787 DLG7	NM_014750	NT_026437	Hs.77695	AB076695 m; BC010658 m; BC016276 m; D13633 m; none p	DLG1; HURP; KIAA0008
PTTG1	= I_958208	GP:CAA11683.1	9232 PTTG1	NM_004219	NT_023133	Hs.252587	AF167560 g; AF167564 g; AF062649 m; AF075242 m; AF095287 m; AJ223953 m; BC026003 m	EAP1; PTTG; HPTTG; TUTR1; SECURIN
BIRC5	= I_960986	SP:Q15392	332 BIRC5	NM_001168	NT_010641	Hs.1578	U75285 g; AB028869 m; AF077350 m; BC000784 m; BC008718 m; none p	API4; EPR-1; SURVIVIN
PTTG3	= I_929699	GP:AAC64411.1	26255 PTTG3	NM_021000	NT_008183	Hs.350968	AF200720 g; AF095289 m	-
HMMR	= I_957819	SP:O75330	3161 HMMR	NM_012484; NM_012485	NT_023133	Hs.72550	AF032862 m; U29343 m; none p; D17297 u	RHAMM
PTTG2	= I_957769	GP:AAC64410.1	10744 PTTG2	NM_006607	NT_016297	Hs.350966	AF116538 g; AF200719 g; AF095288 m	-

TABLE 2B - Part 2

Gene Name	Description	Omim	Location	Pfam	COG5560	Ontologie
USP16	=ubiquitin specific protease 16	604735	21q22.11	Ubiquitin C-terminal hydrolase [Posttranslational modification, protein, turnover, chaperones]		GO:0004843;G O:0004197;GO :0007049;GO: 0006511;GO:0 005737;GO:00 04221;GO:001 6787 molecular function ubiquitin- specific protease activity;molecular function cysteine- type endopeptidase activity;biological process cell cycle;biological process ubiquitin- dependent protein catabolism;cellular component cytoplasm; molecular function ubiquitin C-terminal hydrolase activity;molecular function hydrolase activity
CDC47	=cell division cycle associated 7	606916	2q31			
CKS1	=CDC28 protein kinase regulatory subunit 1B	116900	1q21.2	Cyclin-dependent kinase regulatory subunit	pfam01111	molecular function cyclin- dependent protein kinase activity;biological process regulation of CDK activity;biological process cytokinesis molecular function DNA binding biological process mitotic checkpoint;biologica l process cell cycle;biological process mitosis;cell ular component kinetochor
THAP1	= THAP1	605295	8p11.21	THAP domain	pfam05485	GO:0003677
MAD2L1	=MAD2 mitotic arrest deficient-like 1 (yeast)	601467	4q27	HORMA domain. The HORMA (for Hop1p, Rev7p and MAD2) domain has been suggested to recognise chromatin states that result from DNA adducts,	pfam02301	GO:0007093;G O:0007049;GO :0007067;GO: 0005699;GO:0 005634

Gene Name	Description	Omin	Location	Pfam	Ontologie
KIAA0008	=discs, large homolog 7 (Drosophila)	605584	14q22.2	double stranded breaks or non-attachment to the spindle and acts as an adaptor that recruits other proteins. MAD2 is Guanylate-kinase-associated protein (GKAP) protein	e;cellular component nucleus
PTTG1	=pituitary tumor-transforming 1	604147	5q35.1		molecular function GO:0005554;G function molecular_f O:0000004;GO unction :0007267;GO: unknown;biological 0008372 process biological_p rocess unknown;biological process cell-cell signaling;cellular component cellular_c component unknown molecular GO:0003700;G function transcripti O:0007283;GO on factor :0007048;GO: activity;biological 0006366;GO:0 process spermatogene 005737;GO:00 sis;biological 05634 process oncogenesis; biological process transcriptio n from Pol II promoter;cellular component cytoplasm; cellular component nucleus molecular GO:0008189;G function apoptosis O:0000008;GO inhibitor :0006916;GO: activity;biological 0007048;GO:0 process G2/M 005876 transition of mitotic cell cycle;biological process anti- apoptosis;biological process oncogenesis; cellular
BIRC5	=baculoviral IAP repeat-containing 5 (survivin)	603352	17q25	Baculoviral inhibition of apoptosis protein repeat	

Gene Name	Description	Omin	Location	Pfam	Ontologie
PTTG3	=pituitary tumor- transforming 3	605127	8q13.1	-	component spindle microtubule
HMMR	=hyaluronan- mediated motility receptor (RHAMM)	600936	5q33.2-qter	Chromosome segregation ATPases (Cell division and chromosome partitioning);Chrom osome segregation ATPases (Cell division and chromosome partitioning)	COG1196;COG 1196 molecular function hyaluronic acid binding;biological process cell motility
PTTG2	=pituitary tumor- transforming 2	604231	4p12	-	GO:0005540;G O:0006928

TABLE 2B - Part 3

Gene Name	PubMed	GoldenPath hg16 (7/2003)	Goldenpath (oligos)	Acembly
USP16	= 12477932;10830953;10077596;9827704	chr21:29318881-29348681 + 21q21.3	chr21:29348613-	-
CDCA7	= 11598121	chr2:174422091-174436263 + 2q31.1	29348672 + chr2:174197514-	-
CKS1	= 12473461;8697818;8601310;2227411	chr1:152164005-152168514 + 1q22	174197573 + chr10:30137616-	-
THAP1	= 12477932	chr8:42709181-42715836 - 8p11.21	30137675 + chr8:42432944-	-
MAD2L1	= 12477932;12351790;11912137;11907259;10366450;9615237;9546394;9345911;8824189	chr4:121439410-121446782 - 4q27	42433003 - chr4:121374550-	-
KIAA0008	= 12527899;11543626;9179496;7584028;7584026	chr14:53604888-53648437 - 14q22.3	121374609 - chr14:53605154-	DIG7.b
PTTG1	= 12727994;12590639;12444553;12403781;12355087;12324572;12213878;12194817;10580151;10411507;10393434;10022450;9925941;9915854;9892021;9811450	chr5:159829759-159836640 + 5q33.3	53605213 - chr5:159791187-	-
BIRC5	= 12885482;12833149;12805209;12794243;12773388;12709681;12654446;12643601;12609713;12569609;12556969;12517802;12510154;12419797;12393476;12388702;12374680;12363043;12235242;12174930;12168867;12143224;12133447;12119561;12115583;12085263;12073047;11925104;11888845;11877677;11875736;11861764;11844831;11821157;11773702;11728454;11712083;11084331;9859993;9556606;9256286;8106347;7947793	chr17:76807471-76817900 + 17q25.3	chr17:76681795- 76681854 +	-
PTTG3	= 10806349	chr8:67729592-67730201 - 8q13.2	chr8:67402834- 67402893 -	-
HMMR	= 12712331;12225794;11716065;9601098;8890751;8595891	chr5:162868557-162899840 + 5q34	chr5:162854380- 162854439 +	-
PTTG2	= 10806349;10084610	chr4:37859235-37859811 + 4p14	chr4:37796987- 37797046 +	PTTG2

TABLE 2B - Part 4

Gene Name	Oligo Agilent	SEQ ID NO:
USP16	= GTACTTTGTGTTTAAATATATCTGGGTGATGGATCACAACACATCAATAAACTGACTTACC	519
CDC47	= ATTTACTTGCATATGTAAACCATTTGCTGTGCCATTCAATGTTGATGCATAATTGGACCT	520
CKS1	= AGATGGAGGAGCATCTGAGTTTGAGACCATGGCTGTACAGGGATCATGTAACCTTGCT	521
THAP1	= TGGAGATTTAAACACTGAGGTTTCTGTTCAACTGTGAGTTCTGTTCTTTGTGAGAAAT	522
MAD2L1	= TGTACCTGAAAATGGGAAGAGTCGGGACCACAGTTTATTACCAATTTCTGAGGAAGTCCG	523
KIAA0008	= ATCCATTACTCAGCTGGAGAGGAGACATCAAGAACATGCCAGACACATTTCTTTGGTG	524
PTTG1	= CTGGATGTTGAATTGCCACCTGTTGCTGTGACATAGATTTAAATTTCTTAGTGCTTC	525
BIRC5	= CTGGAAACCTCTGGAGGTCATCTCGGTGTTCTCGAGAAATAAAAGCCTGTCAATTTCAA	526
PTTG3	= TGTTCAGTCTCCTTTAAGCATCTGTTGACCCCTGGATGTTGAATGCCACCTGTTGCT	527
HMMR	= ACTATTCTTCAGAGTTTGTCAATATCTGCTTGTCTCATCTGATGTCTACTCAGCATTTGA	528
PTTG2	= AGACTGTTAAACAAAAGTTCTGTTCTCCTGCCTCAGATGACGCCCTATCCAGAAATAGAAA	529

Out of 17000 genes examined in these microarray experiments, we identified 23 candidate THAP1-target genes that are downregulated in THAP1-overexpressing cells. One of the genes identified corresponds to THAP1 itself (FLJ10477), suggesting auto-regulation. Nine genes correspond to predicted proteins with unknown functions. Strikingly, at least 10 of the remaining  
 5 13 genes downregulated by THAP1 (see Table 2A) correspond to proteins previously linked to cell cycle/cell proliferation (CKS1, Survivin, PTTG1/Securin, PTTG2/Securin2, PTTG3/Securin3, MAD2L1, USP16, HMMR, KIAA0008, CDCA7). Many of these genes share common characteristics.

1) role in mitosis/chromosome segregation: **Survivin** (polypeptide sequence SEQ ID NO: 343, nucleotide sequence SEQ ID NO: 344) (Li et al. (1998) Nature 396:580-584; Li et al. (1999) Nature Cell Biol 1:461-466; Lens et al. (2003) EMBO J 22:2934-2947), **PTTG1/Securin** (polypeptide sequence SEQ ID NO: 345, nucleotide sequence SEQ ID NO: 346) (Zou et al. (1999) Science 285:418-422; Wang et al. (2001) Mol Endocrinol 15:1870-1879), **CKS1** (polypeptide sequence SEQ ID NO: 347, nucleotide sequence SEQ ID NO: 348) (Pines (1996) Curr Biol 6:1399-1402; Hixon et al. (2000) J Biol Chem 275:40434-40442), **MAD2L1** (polypeptide sequence SEQ ID NO: 349, nucleotide sequence SEQ ID NO: 350) (Dobles et al. (2000) Cell 101:635-645; Michel et al. (2001) Nature 409:355-359), **USP16/Ubp-M** (polypeptide sequence SEQ ID NO: 351, nucleotide sequence SEQ ID NO: 352) (Cai et al. (1999) PNAS 96:2828-2833), **HMMR/RHAMM** (polypeptide sequence of isoform A, SEQ ID NO: 353, nucleotide sequence of transcript variant 1, SEQ ID NO: 354) (polypeptide sequence (gi/7108351) SEQ ID NO: 365, nucleotide sequence of transcript variant 2, SEQ ID NO: 366) (Maxwell et al. (2003) Mol Biol Cell 14:2262-2276; Tolg et al. (2003) Oncogene 22:6873-6882), **KIAA0008/HURP** (polypeptide sequence SEQ ID NO: 355, nucleotide sequence SEQ ID NO: 356) (Tsou et al. (2003) Oncogene 22:298-307);

2) specific mRNA expression in S/G2-M: **CKS1** (Richardson et al. (1990) Genes Dev 4:1332-1344), **Survivin** (Li et al. (1998) Nature 396:580-584), **PTTG1/Securin** (Zou et al. (1999) Science 285:418-422; Yu et al. (2000) Mol Endocrinol 14:1137-1146), **KIAA0008/HURP** (Bassal et al. (2001) Genomics 77:5-7; Tsou et al. (2003) Oncogene 22:298-307);

3) upregulation in human tumors: **CKS1** (Inui et al. (2003) BBRC 303:978-984), **Survivin** (Ambrosini et al. (1997) Nature Med 3:917-921), **PTTG1/Securin** (Heaney et al. (2000) Lancet 355:716-719; Zou et al. (1999) Science 285:418-422), **PTTG2/Securin2** (polypeptide sequence SEQ ID NO: 357, nucleotide sequence SEQ ID NO: 358) (Chen et al. (2000) Gene 248:41-50), **PTTG3/Securin3** (polypeptide sequence SEQ ID NO: 359,



- nucleotide sequence SEQ ID NO: 360) (Chen et al. (2000) Gene 248:41-50), **HMMR/RHAMM** (Tolg et al. (2003) Oncogene 22:6873-6882), **KIAA0008/HURP** (Bassal et al. (2001) Genomics 77:5-7; Tsou et al. (2003) Oncogene 22:298-307), **CDCA7/JPO1** (polypeptide sequence of variant 1, SEQ ID NO: 361, nucleotide sequence of variant 1, SEQ ID NO: 362; polypeptide sequence of isoform 2, SEQ ID NO: 363, nucleotide sequence of transcript variant 2, SEQ ID NO: 364) (Prescott et al. (2001) J Biol Chem 276:48276-48284);
- 4) negative regulation by the p53 tumor suppressor: **Survivin** (Hoffman et al. (2002) J Biol Chem 277:3247-3257; Mirza et al. (2002) Oncogene 21:2613-2622), **PTTG1/Securin** (Zhou et al. (2003) J Biol Chem 278:462-470);
- 5) stimulation of angiogenesis: **Survivin** (O'Connor et al. (2000) Am J Path 156:393-398; Papapetropoulos et al. (2000) J Biol Chem 275:9102-9105; Mesri et al. (2001) Am J Path 158:1757-1765), **PTTG1/Securin** (Ishikawa et al. (2001) J Clin Endocrinol Metab 86:867-874; McCabe et al. (2002) J Clin Endocrinol Metab 87:4238-4244).

In addition, survivin has been shown to be a critical anti-apoptotic factor at the interface between cell cycle/mitosis and apoptosis (Li et al. (1998) Nature 396:580-584; Li et al. (1999) Nature Cell Biol 1:461-466), which plays an important role in the control of endothelial cell apoptosis (O'Connor et al. (2000) Am J Path 156:393-398; Papapetropoulos et al. (2000) J Biol Chem 275:9102-9105; Mesri et al. (2001) Am J Path 158:1757-1765). Downregulation of survivin expression by THAP1 may therefore contribute to its pro-apoptotic activity (see Example 10). Simultaneous downregulation by THAP1 of all these genes critical for cell cycle/cell proliferation and/or apoptosis (CKS1, Survivin, PTTG1/Securin, MAD2L1, USP16, HMMR), is expected to result in cell cycle block and inhibition of cell proliferation. Accordingly, we found that overexpression of THAP1 in primary human endothelial cells or human U2OS osteosarcoma cancer cells resulted in inhibition of cell proliferation after a few days, followed by apoptosis.

#### EXAMPLE 46

##### THAP1 responsive elements in cell cycle-specific THAP1 target genes

We searched the promoters of the THAP1-target genes for the presence of THAP1-responsive elements. This analysis allowed us to identify candidate DR-5 or THRE motifs that may mediate direct binding of THAP1 to the promoters of its target genes. A candidate DR5-type THAP1 responsive element (GGGCAAnnnnnGGGCAC) (SEQ ID NO: 316) located in the antisense orientation close to the AUG codon of the human *survivin/BIRC5* gene is shown in Figure 30. A candidate THRE-type THAP1 responsive element (AGTGTGGGCAT) (SEQ ID NO: 318)

located in the antisense orientation near the TATA box of the *Ubiquitin specific protease 16* gene is shown in Figure 31.

## EXAMPLE 47

Chemokine SLC/CCL21 modulates transcription of  
cell-cycle specific genes in a THAP1-dependent manner

5 To examine the effects of the nuclear SLC/THAP1 complex on global expression profiles in human primary endothelial cells (HUVEC), we performed microarray experiments with cells successively transduced with SLC/CCL21 chemokine and THAP1 (SLC/THAP) retrovirus expression vectors or control cells transduced with MCS/THAP1 or SLC/MCS retrovirus  
10 expression vectors. A hierarchical cluster analysis was performed based on similarity of expression patterns of genes.

Table 3A lists the genes downregulated by the SLC/THAP1 complex in human primary endothelial cells, as revealed by the above-describe microarray experiments. For each gene, the fold changes, p values and signal intensities obtained in the three microarray experiments are  
15 indicated.

Table 3B lists database accession numbers and SEQ ID NOs. for each gene and corresponding polypeptide listed in Table 3A.

TABLE 3A

Sequence Name(s)	Experiment Name Sequence Description	MCS/THAPvsMCS				SLC/THAPvsMCS				SLC/MCSvsMCS			
		Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2
CKS1	CDC28 protein kinase 1, binds and regulates CDK2-cyclin A complexes, similar to S. pombe p13suc1, required for SCF dependent degradation of p27	-2.01	2.20E-24	27653	13620.7	-2.4	0	23175.76	9619.37	-1.22	8.53E-04	23654.3	19415.75
PTTG2	Pituitary tumor transforming gene 2, protein with strong similarity to human PTTG1, which is a proto-oncoprotein with PXXP motifs that may bind to SH3 domains, suggesting roles in intracellular signal transduction and tumor-specific pathogenesis	-1.56	8.44E-07	4727.5	3064.92	-2.25	2.65E-10	4390.91	1952.05	-1.22	0.06	3336.21	2799.64
CDKN3	Cyclin-dependent kinase inhibitor 3 (cyclin-dependent kinase interactor 1), a tyrosine-serine phosphatase that interacts with cyclin-dependent kinases and inhibits progression through the cell cycle, dephosphorylates CDK2 monomer on Thr160	-1.53	2.08E-14	4984.3	3267.56	-2.05	9.79E-16	4124.82	1998.74	-1.24	2.19E-05	4168.48	3367.27
BUB1	Budding uninhibited by benzimidazoles 1 homolog, a spindle assembly checkpoint protein that may sense kinetochore tension; mutations are associated with lung cancer, adult T cell leukemia, and chromosome instability in colorectal cancer cell lines	-1.38	1.44E-04	1623.9	1195.58	-2.05	2.71E-08	1528.46	746.68	-1.07	0.41	1162.03	1087.69
HMMR	Hyaluronan mediated motility receptor, binds hyaluronan and is important for cell motility, binds microtubules and microfilaments intracellularly, may also be involved in cell proliferation;	-1.69	9.91E-06	2255.3	1286.07	-1.94	1.76E-15	1987.08	1019.59	-1.25	0.02	1779.65	1405.32

Sequence Name(s)	Experiment Name Sequence Description	MCS/THAPvsMCS				SLC/THAPvsMCS				SLC/MCSvsMCS			
		Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2
UISNRNPBP	gene mutations may contribute to colon cancer development U1-snRNP binding protein homolog (70kD), arginine rich basic protein similar to U1 70K splicing factor (SNRP70); two alternative forms identified which differ in the 5 untranslated region	-1.92	4.00E-05	904.14	461.82	-1.89	5.93E-08	799.93	417.53	-1.17	0.1	852.82	729.99
HIF5	H1 histone family member 5, a linker histone involved in compaction of nucleosomes into high-order chromatin structures	-1.39	0.15	445.51	314.07	-1.89	1.87E-04	411.96	216.91	-1.14	0.51	310.27	269.82
PTTG3	Pituitary tumor-transforming 3, a protein that may be associated with tumorigenesis	-1.54	2.41E-24	20058	13058.5	-1.89	2.23E-21	17289.26	9167.28	-1.16	7.79E-04	14947.8	12937.42
TOPK	PDZ-binding kinase, a serine-threonine kinase active during mitosis which binds PDZ domain-containing proteins, also activates p38 MAP kinase, may be involved in cell cycle regulation, lymphoid cell activation, and spermatogenesis	-1.43	1.43E-06	3873.8	2735.7	-1.88	3.71E-19	3428.27	1822.17	-1.13	0.08	3004.03	2669.67
PTTG1	Pituitary tumor-transforming 1 (securin), a transcriptional activator that promotes cell proliferation and angiogenesis, involved in sister chromatid separation and euploidy maintenance, overexpression promotes cellular transformation and tumorigenesis	-1.53	3.65E-17	20405	13309	-1.87	3.45E-30	19057.67	10181.17	-1.19	1.96E-03	16411.9	13758.9
HIF3	H1 histone family member 3, involved in compaction of DNA into nucleosomes and into high-order chromatin structures	-1.28	2.85E-03	491.79	382.35	-1.83	9.37E-04	541.92	290.18	-1.27	0.05	430.9	339.32
KIAA0008	Member of the guanylate-kinase-associated protein (GKAP)	-1.54	1.36E-24	5911.6	3848.34	-1.81	1.81E-36	5327.3	2939.49	-1.14	0.01	4841.06	4268.61

Sequence Name(s)	Experiment Name Sequence Description	MCS/THAPvsMCS			SLC/THAPvsMCS			SLC/MCSvsMCS		
		Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2	Intensity 2
CDCA7	protein family									
BIRC5	c-myc target IPO1 Survivin, a member of the inhibitor of apoptosis protein family that is involved in G2-M transition and exit of the mitotic cell cycle; may play a role in oncogenesis	-1.6	1.53E-07	2114.2	1321.11	-1.8	2.12E-06	1704.58	946.32	1440.39
		-1.55	2.90E-08	17927	11605.8	-1.79	2.66E-13	18270.21	10211.76	11869.49
CNAP1	Protein with low similarity to yeast LOC7, which is required for sister chromatid separation and segregation; a component of the condensin complex which includes CAP-E and CAP-C (1053947)	-1.39	6.85E-03	1441.4	1022.72	-1.78	9.22E-29	1332.49	747.25	1056.96
FLJ10477	THAP1	-1.92	5.96E-17	1510.6	791.48	-1.75	6.02E-08	1321.01	757.89	1150.59
USP16	Ubiquitin specific protease 16, deubiquitinates histone H2A and H2B, may play a role in deubiquitination of proteins involved in the condensation of mitotic chromosomes and may deubiquitinate histone H2A during apoptosis	-1.66	6.49E-10	1224.1	739.65	-1.75	1.49E-19	1722.79	979.19	742.12
MAD2L1	Mitotic arrest deficient 2 yeast homolog-like 1, essential for timing of anaphase onset, involved in detection of kinetochore attachment as part of the mitotic spindle checkpoint; genetic variants may be linked to breast cancer	-1.69	1.36E-05	3694.8	2191.56	-1.75	3.68E-06	3934.21	2248.14	2479.62
CDCA1	Member of the NuF2 family, which are components of mitotic spindles	-1.42	1.07E-15	1735.6	1219.5	-1.74	2.42E-21	1247.39	713.27	1071.52
BUB1B	BUB1 budding uninhibited by benzimidazoles 1 homolog beta, a protein kinase that acts in the mitotic spindle checkpoint and	-1.21	5.65E-03	1327.8	1102.95	-1.74	1.07E-16	1177.27	675.06	1084.43

Experiment Name		MCS/THAPvsMCS				SLC/THAPvsMCS				SLC/MCSvsMCS			
Sequence Name(s)	Sequence Description	Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2
	inhibits anaphase-promoting complex activation; genetic mutations/inactivation is associated with leukemia and colorectal cancers												
DUT	dUTP pyrophosphatase, maintains dUTP at low levels to prevent misincorporation into DNA during replication, mediates resistance to 5-fluorouracil, may regulate peroxisome proliferation; alternative splicing generates nuclear and mitochondrial forms	-1.33	1.25E-04	9787.6	7429.53	-1.72	4.41E-08	8255.49	4778.82	-1.18	0.06	7430.5	6384.26
KNSL7	Kinesin-like 7 (kinesin-like protein 2), a putative motor protein that may modulate mitotic progression, interacts with the forkhead-associated domain of pK1-67 (MK167), which is a cell proliferation marker protein	-1.41	0.03	700.06	484.71	-1.72	1.03E-05	746.58	431.14	1.03	0.89	560.37	570.29
KNSL1	Kinesin-like 1, a microtubule-associated kinesin motor that functions in mitotic spindle formation and centrosome separation, and acts antagonistically with the minus-end directed kinesin KNLSL2, localized to spindle is regulated by CDC2 phosphorylation	-1.28	7.48E-05	1014.4	800.59	-1.71	2.35E-14	753.82	440.83	-1.16	0.01	658.34	565.96
CCNB2	Cyclin B2, a CDC2 kinase-associated cyclin that is involved in Golgi apparatus disassembly, may function in p53 (TP53) - mediated cell cycle arrest at the G2/M transition, may mediate cell cycle arrest by linking CDC2 with TGFbeta type II	-1.42	6.32E-06	9919.2	7032.51	-1.7	5.03E-10	8702.06	5119.46	-1.09	0.18	7515.61	6907.19

Sequence Name(s)	Experiment Name Sequence Description	MCS/THAPv6MCS				SLC/THAPv6MCS				SLC/MCSv6MCS			
		Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2
CHEK2	receptor (TGFBR2) CHK2 checkpoint homolog, protein kinase involved in DNA damage response and cell cycle arrest, phosphorylated by ataxia telangiectasia mutated kinase (ATM), phosphorylates p53 (TP53) and mediates BRCA1 function; downregulated in some breast cancers	-1.29	1.03E-04	975.21	754.54	-1.7	3.18E-13	792.26	467.02	1.01	0.93	653.15	656.44
CDC2	Cell division cycle 2, cyclin-dependent protein kinase that binds B-type cyclins, regulates entry of mitosis and G2 to M-phase transition, promotes cell proliferation; implicated in Alzheimers disease through phosphorylation of amyloid beta and nucleolin	-1.3	2.62E-04	4542.3	3495.79	-1.7	1.07E-06	4897.33	2869.04	-1.14	0.11	4164.17	3682.1
CCNB1	Cyclin B1, regulatory subunit of the CCNB1 - CDC2 maturation-promoting factor complex that mediates G2-M phase transition, plays a role in radiation-induced apoptosis, overexpression induces tetraploidy	-1.44	3.93E-07	11740	8158.53	-1.69	1.04E-23	10499.04	6193.63	-1.15	0.05	9515.79	8323.01
KNL4	Kinesin-like 4, a DNA and microtubule-binding protein, associates with mitotic chromosomes and is enriched in the kinetochore during anaphase, involved in generating the away-from-the-pole force necessary for chromosome oscillation during mitosis	-1.25	1.14E-05	3350.3	2685.52	-1.68	5.65E-19	2817.57	1684.53	-1.07	0.15	2771.76	2600.21
LOC51053	Geminin, an inhibitor of DNA replication that may regulate the DNA replication process by inhibiting inappropriate firing of the replication origin through	-1.43	2.78E-09	6058.4	4204.6	-1.67	6.44E-32	5448.87	3265.67	-1.22	1.12E-03	4644.96	3775.2

Sequence Name(s)	Experiment Name Sequence Description	MCS/THAPvsMCS				SLC/THAPvsMCS				SLC/MCSvsMCS			
		Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2
CKS2	binding to CDT1 Protein that binds to CDC2/CDC28 protein kinase, regulates CDK-cyclin complexes; similar to S. pombe p13suc1	-1.14	0.26	17790	15646.8	-1.67	1.20E-14	19053.8	11409.25	-1.14	0.04	16638.8	14544.77
PRC1	Protein regulator of cytokinesis 1, associates with the mitotic spindle and is required for cytokinesis but not nuclear division, may function in spindle elongation, a substrate for phosphorylation by many cyclin dependent kinases	-1.36	1.04E-12	6118	4493.41	-1.65	2.28E-25	5565.83	3378.6	-1.1	0.03	4881.13	4455.85
TYMS	Thymidylate synthetase, catalyzes the reductive methylation of dUMP to dTMP with concomitant conversion of 5,10-methylenetetrahydrofolate to dihydrofolate	-1.13	0.28	14694	12991.1	-1.65	8.90E-08	16860.03	10195.4	-1.14	0.13	14876.8	12943.79
E2-EPF	Keratinocyte ubiquitin carrier protein, which is required for ubiquitin-protein conjugation, links ubiquitin with a thiol ester linkage in a ubiquitin activating enzyme-dependent manner, may be associated with endemic pemphigus foliaceus (EPF)	-1.35	2.48E-04	11005	8132.69	-1.65	1.20E-29	10039.67	6073.81	-1.11	0.27	9649.67	8699.99
SMARCD1	SWI-SNF related matrix associated actin dependent regulator of chromatin subfamily d member 1, part of complexes implicated in regulation of transcription by remodeling chromatin and involved in regulation of fetal to adult globin gene switch	-1.23	0.26	13768	110.98	-1.64	6.97E-07	196.79	119.78	-1.05	0.73	176.74	169.51
HSPC150	Protein with high similarity to S. cerevisiae Ubc13p, which is a ubiquitin-conjugating (E2)	-1.32	1.33E-07	5559.4	4238.57	-1.63	2.08E-19	5064.71	3107.47	-1.18	9.95E-03	4533.44	3853.02



Experiment Name		MCS/THAPvsMCS				SLC/THAPvsMCS				SLC/MCSvsMCS			
Sequence Name(s)	Sequence Description	Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2
	enzyme involved in the S. cerevisiae Rad6 dependent post-replicative repair pathway, member of the ubiquitin-conjugating enzyme (E2) family												
LSM5	U6 snRNA-associated Sm-like protein, a putative RNA-binding protein that forms a doughnut-shaped U6 snRNA-containing complex with other Sm-like proteins, may play a role in U4/U6 snRNP formation	-1.21	2.14E-03	6921.7	5751.08	-1.63	3.82E-05	5718.65	3504.33	-1.24	8.24E-04	5998.6	4857.88
SF3A3	Spliceosome-associated protein 61, a subunit of the heterotrimeric splicing factor SF3a, involved in the formation of the 17S U2 snRNP and assembly of the prespliceosome, contains a C2H2 zinc finger	-1.13	0.53	706.9	628.91	-1.63	9.67E-05	1828.86	508.89	-1.13	0.61	600.6	522.95
HEC	Highly expressed in cancer, a nuclear protein that localizes to the centromere during M phase, inhibits proteolysis of M phase cyclin B, may be involved in chromosome segregation and M phase progression, high level expression is observed in cancer cells	-1.36	7.33E-05	5357.5	3893.36	-1.62	4.57E-11	4750.6	2931.09	-1.11	0.17	4382.36	3946.21
FEN1	Flap structure specific endonuclease 1, multifunctional endonuclease and exonuclease that has roles in DNA replication and repair; interaction with PCNA stimulates nuclease activity, may be involved in trinucleotide repeat expansion-related diseases	-1.44	9.40E-05	2312.8	1586.37	-1.61	5.76E-23	2088.97	1305.57	-1.16	0.03	1874.86	1614.81
ZWINT	ZW10 interactor, a kinetochore protein with an extended coiled coil domain, interacts with	-1.37	1.57E-04	5167.1	3790.55	-1.61	5.03E-08	4733.23	2956.31	-1.08	0.34	4183.75	3883

Experiment Name		MCS/THAPvsMCS				SLC/THAPvsMCS				SLC/MCSvsMCS			
Sequence Name(s)	Sequence Description	Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2
	ZW10 and may target it to the kinetochore at prometaphase, may have a role in centromere function												
DTYMK	Deoxythymidylate kinase (dTMP kinase), phosphorylates dTTP to dTDP in dTTP biosynthesis, activity and transcript abundance peak in S phase, rate limiting for activating zidovudine (AZT)	-1.3	2.06E-04	3500.6	2732.79	-1.6	2.98E-07	3297.46	2039.57	-1.02	0.83	2616.54	2554.15
RADI	RADI homolog (S. pombe), a 3 to 5 exonuclease with predicted roles in DNA damage-activated mitotic and meiotic cell cycle checkpoints, involved in DNA repair and the DNA damage response	-1.13	0.17	582.29	515.05	-1.6	3.77E-05	502.15	310.68	-1.23	0.08	579.61	469.97
TACC3	Transforming acidic coiled-coil containing protein 3, microtubule-binding protein that may regulate cell growth and microtubule nucleation; corresponding chromosomal region is disrupted in multiple myeloma	-1.33	3.44E-06	963.58	727.29	-1.6	1.57E-23	948.96	594.38	-1.23	0.04	822.02	665.45
GTSE1	G-2 and S-phase expressed 1, a cell cycle regulated, microtubule-associated protein, ectopic expression inhibits progression through the G2 to M transition of the cell cycle	-1.18	0.29	566.43	478.95	-1.6	1.30E-13	517.38	324.08	-1.18	0.33	451.63	381.19
RRM2	Ribonucleotide reductase subunit M2, associates with RRM1 to yield an enzyme that reduces ribonucleotides to deoxyribonucleotides, a rate limiting enzyme for DNA synthesis; overexpression is associated with hydroxyurea	-1.48	5.34E-11	18203	12337.2	-1.59	5.54E-11	16556.89	10390.69	-1.17	0.06	15290.2	13175.31

Sequence Name(s)	Experiment Name	MCS/THAPvsMCS				SLC/THAPvsMCS				SLC/MCSvsMCS			
		Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2
RAB6KIFL	resistance RAB6 interacting kinesin-like (rabkinesin 6), a putative kinesin-like microtubule motor protein, plays a role in cytokinesis and anaphase B, may play a role in vesicular transport	-1.43	1.22E-04	1113.8	789.17	-1.58	1.33E-18	1025.6	646.53	-1.01	0.91	927.98	924.07
MCM7	Minichromosome maintenance deficient 7, forms a DNA dependent ATP dependent DNA helicase with MCM4 and MCM6, dissociates from replicated chromatin, likely to be involved in DNA replication	-1.29	6.78E-03	15180	11824.5	-1.58	7.32E-06	13991.23	8763.58	-1.08	0.41	12644.5	11783.21
ANLN	Aniline, an actin binding protein that interacts with cleavage furrow proteins such as septins and may play a role in cytokinesis	-1.43	1.31E-12	5924.3	4158.55	-1.57	3.30E-20	5307.32	3390.98	-1.07	0.12	4790.36	4474.32
DTYMK	Deoxythymidylate kinase (dTMP kinase), phosphorylates dTMP to dTDP in dTTP biosynthesis, activity and transcript abundance peak in S phase, rate limiting for activating zidovudine (AZT)	-1.35	1.33E-05	19500	14395.9	-1.57	2.31E-14	17065.82	10894.34	-1.15	0.02	16770.7	14607.91
SNRPG	Sm core protein G, a component of spliceosomal snRNPs that is involved in snRNP formation; a target of antibodies in patients with the autoimmune disease, systemic lupus erythematosus	-1.3	8.82E-06	50664	38792	-1.57	1.72E-17	42579.99	27124.5	-1.23	2.59E-05	40762.1	33037.4
CSE1L	CSE1 chromosome segregation 1-like (yeast), importin-alpha nuclear export receptor, functions in toxin and TNF resistance and apoptosis, may regulate cell proliferation; corresponding gene is amplified in breast and colon carcinoma	-1.33	1.40E-11	17651	13301.8	-1.56	3.02E-13	15607.94	9962.65	-1.17	7.55E-04	14700.2	12604.67

Sequence Name(s)	Experiment Name	MCS/THAPvsMCS				SLC/THAPvsMCS				SLC/MCSvsMCS			
		Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2
TK1	cell lines Thymidine kinase 1, a cytosolic form of the enzyme that synthesizes thymidylate for DNA synthesis, and activates nucleoside analog antiviral and anticancer drugs	-1.34	1.55E-04	31745	23955.2	-1.55	1.65E-07	28002.74	18016.22	-1.25	0.03	25068	20107.36
RRM1	Ribonucleotide reductase M1 subunit, may associate with either RRM2 (S-phase induced) or p53R2 (UV-induced) to yield a functional complex that reduces ribonucleotides to deoxyribonucleotides, a rate limiting step for DNA synthesis	-1.36	6.79E-04	12340	9146.53	-1.55	1.80E-04	11613.5	7473.84	-1.05	0.6	11176	10706.73
ASK	Activator of S phase kinase, binds to and activates kinase activity of CDC7 (CDC7L1), which is required for the initiation of DNA replication at the G1/S transition	-1.26	5.54E-06	1026.1	814.96	-1.55	2.92E-21	930.4	600.22	-1.19	0.04	860.08	734
SNRPA1	Small nuclear ribonucleoprotein polypeptide A, component of the U2 snRNP particle which is a required constituent of the spliceosome	-1.45	3.55E-12	17423	12054.6	-1.54	6.76E-19	15593.29	10109.06	-1.35	1.24E-08	14000.5	10368.48
TUBA4	Member of the tubulin-FzA family, which are involved in polymer formation, has strong similarity to a region of mouse Tuba6, which is a structural protein that polymerizes to form microtubules	-1.2	0.02	20738	17208.3	-1.54	5.19E-14	17826.68	11580.37	-1.19	0.01	15732.1	13207.69
BCL2A1	BCL2-related protein, a member of the Bcl-2 family of apoptosis regulators; inhibits apoptosis, promotes tumorigenesis, and may play a protective role during inflammation	-1.11	0.08	4671.9	4226.81	-1.53	2.03E-08	4032.4	2626.38	-1.36	6.38E-05	3745.61	2748.61

Sequence Name(s)	Experiment Name Sequence Description	MCS/THAPvsMCS			SLC/THAPvsMCS			SLC/MCSvsMCS		
		Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2	Intensity 2
ERH	Enhancer of rudimentary (Drosophila) homolog, may function in pyrimidine metabolism and cell cycle control	-1.13	0.32	29487	25833.5	-1.53	2.03E-09	25423.05	16680.14	21682.35
TTK	Dual specificity serine/threonine and tyrosine kinase, may play a role in IL2-induced cell cycle progression of T cells, may play a role in cartilage homeostasis modulated by TNF alpha (TNF)	-1.23	0.05	640.83	528.12	-1.53	9.09E-05	580.48	378.02	443.64
KNSL6	Mitotic centromere-associated kinesin (Kinesin-like 6), a member of the kinesin family of microtubule-associated motor proteins involved in mitosis; interacts with kinetochore protein CENPH	-1.21	1.82E-05	1994.9	1647.2	-1.52	1.30E-09	1970.84	1292.04	1670.11
CDC45L	Cell division cycle 45 like, associates with ORC2L, MCM7, and POLA2, predicted to be involved in the initiation of DNA replication; corresponding gene is located in a chromosomal region frequently deleted in DiGeorge syndrome	-1.2	0.05	3960.3	3331.41	-1.52	1.45E-06	3616.42	2382.26	2983.49
HIF4	H1 histone family member 4, involved in compaction of DNA into nucleosomes and high-order chromatin structures, may maintain a low methylation state in CpG-rich DNA and linker DNA, may play a role in DNA accessibility during apoptotic DNA fragmentation	-1.2	0.27	1070.1	901.4	-1.51	1.68E-06	1032.36	703.15	939.89
RAD54B	RAD54B, a member of the SNF2-SWI2 superfamily and ATPase that binds human RAD51 and may have a role in cell growth, mitotic	-1.17	0.05	366.03	308.11	-1.51	3.69E-06	352.53	230.42	268.9

Sequence Name(s)	Experiment Name Sequence Description	MCS/THAPvsMCS				SLC/THAPvsMCS				SLC/MCSvsMCS			
		Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2
HMG2	recombination, and double-strand break repair, associated with primary lymphoma and colon cancer upon gene mutation												
	High mobility group (nonhistone chromosomal) protein 2, binds undamaged DNA, involved with HIV-1 integration and viral replication, reacts with sera of subjects having autoimmune hepatitis disease and juvenile idiopathic arthritis	-1.31	6.21E-04	14713	11260.5	-1.5	1.70E-11	12957.86	8610.37	-1.06	0.42	10763.2	10233.57
CDC20	Cell division cycle 20, seven WD repeat protein that is essential for cell division, interacts with and activates the mitotically phosphorylated form of the anaphase promoting complex, involved in mitotic spindle checkpoint activation	-1.35	2.15E-03	19505	14623.7	-1.5	4.29E-08	16764.25	11109.96	-1.06	0.51	14846.6	14054.45
MCM10	Homolog of Saccharomyces cerevisiae Mcm10p, interacts with replication factors ORC2L, MCM2 and MCM6, may function in initiation of DNA replication	-1.39	2.08E-07	693.96	499.14	-1.48	9.70E-04	578.36	390.08	-1.19	6.92E-03	551.71	461.94
KPNB1	Importin beta (karyopherin beta 1), a subunit of the NLS (nuclear localization signal) receptor complex, binds to the nuclear pore complex and mediates translocation of the importin alpha-NLS complex into the nucleus	-1.28	1.65E-03	2317	1801.94	-1.4	1.16E-05	1978.38	1416.8	-1.01	0.88	1923.44	1909.19
FOXMI	Forkhead box M1, member of the HFN-3/fork head/winged-helix family of transcription factors, has roles in cell proliferation, cell cycle control,	-1.38	8.89E-07	1175.3	853.51	-1.37	2.35E-08	1046.96	762.08	-1.15	0.02	957.6	837.68

Sequence Name(s)	Experiment Name	MCS/THA Pvs MCS				SLC/THA Pvs MCS				SLC/MCS vs MCS			
		Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2	Fold Change	P-Value	Intensity 1	Intensity 2
USP15	and response to oxidative stress Ubiquitin-specific protease 15, a member of the ubiquitin-specific cysteine (thiol) protease family, cleaves ubiquitin from ubiquitin-conjugated protein substrates, may play a role in growth regulation	-1.08	0.32	490.48	453.2	-1.36	4.88E-05	461.58	338.04	-1.14	0.31	456.38	401.43
L_1109838	Protein with strong similarity to dihydrofolate reductase (mouse Dhfr), which catalyses NADPH-dependent reduction of dihydrofolate to tetrahydrofolate, member of the dihydrofolate reductase family	-1.32	0.05	384.36	269.86	-1.36	1.02E-06	295.27	215.01	-1.58	2.34E-04	315.82	198.25
DNMT1	DNA (cytosine-5-)-methyltransferase, may carry out both de novo and maintenance DNA methylation, deregulated expression may promote cellular transformation	-1.17	8.21E-03	8488.9	7236.21	-1.25	9.38E-05	7772.65	6214.13	-1.04	0.53	6919.14	6678.37

TABLE 3B  
(SLCTHAP1)

Gene Name	Agilent	Protein	RefSeq	Unigene	GenBank	Description	Amino Acid SEQ ID NO.	Nucleic Acid SEQ ID NO.
CKS1	I_929087	SP:P33551	NM_001826.1		BC001425.1		367	368
PTTG2	I_957769	GP:AAC64410.1	NM_006607.1	Hs.511755	AF095288.1	pituitary tumor-transforming 2	369	370
CDKN3	I_959891	SP:Q16667	NM_005192.2	Hs.84113	L27711.1	cyclin-dependent kinase inhibitor 3 (CDK2-associated dual specificity phosphatase)		
BUB1	I_965088	SP:O43683	NM_004336.1	Hs.287472	AF046078.1	BUB1 budding uninhibited by benzimidazoles 1 homolog (yeast)	371	372
HMMR	I_957819	SP:O75330	NM_012484.1	Hs.72550	AF032862.1	hyaluronan-mediated motility receptor (RHAMM)	373	374
UISNRNPBP	I_965626	GP:AAA86654.1	NM_007020.1; NM_022717.1	Hs.427552	U44799.1	U1-snRNP binding protein homolog	375	376
HIF5	I_958019	SP:P16401	NM_005322.2		X83509.1		377	378
PTTG3	I_929699	GP:AAC64411.1	NM_021000.1	Hs.521097	AF095289.1	pituitary tumor-transforming 3	381	382
TOPK	I_929157	GP:BAA99576.1	NM_018492.2	Hs.104741	AB027249.1	T-LAK cell-originated protein kinase	383	384
PTTG1	I_958208	GP:CAA11683.1	NM_004219.2	Hs.350966	AJ223953.1	pituitary tumor-transforming 1	385	386
HIF3	I_957891	SP:P16402	NM_005320.1		M60747.1		387	388
KIAA0008	I_959284	SP:Q15398	NM_014750.1	Hs.77695	BC010658.1	discs, large homolog 7 (Drosophila)	389	390
CDCA7	I_928296	GP:AAH27966.1	NM_031942.1	Hs.435733	AK027642.1	cell division cycle associated 7	391	392
BIRC5	I_960986	SP:O15392	NM_001168.1		U75285.1		393	394
CNAP1	I_936441	GP:BAA09930.1	NM_014865.1	Hs.5719	D63880.1	chromosome condensation-related SMC-associated protein 1	395	396
FLJ10477	I_929644	GP:BAA91635.1	NM_018105.1	Hs.7432	AK001339.1	THAP domain containing, apoptosis associated protein 1	397	398
USP16	I_962079	SP:Q9Y5T5	NM_006447.1	Hs.99819	AK023247.1	ubiquitin specific protease 16	399	400
MAD2L1	I_957747	SP:Q13257	NM_002358.2	Hs.79078	BC000356.1	MAD2 mitotic arrest deficient-like 1 (yeast)	401	402
CDCA1	I_942438	GP:BAB59141.1	NM_031423.1	Hs.234545	AF326731.1	cell division cycle associated 1	403	404
BUB1B	I_958935	SP:O60566	NM_001211.2	Hs.36708	AF053306.1	BUB1 budding uninhibited by benzimidazoles 1 homolog beta (yeast)	405	406
							407	408



Gene Name	Agilent	Protein	RefSeq	Unigene	GenBank	Description	Amino Acid SEQ ID NO.	Nucleic Acid SEQ ID NO.
DUT	I_958567	GP:BAB13724.1	NM_001948.1	Hs.367676	U31930.1	dUTP pyrophosphatase	409	410
KNL7	I_954746	GP:BAB03309.1	NM_020242.1	Hs.315051	AB035898.1	kinesin-like 7	411	412
KNL1	I_931999	SP:P52732	NM_004523.2	Hs.8878	U37426.1	kinesin family member 11	413	414
CCNB2	I_959997	SP:O95067	NM_004701.2	Hs.194698	AL080146.1	cyclin B2	415	416
CHEK2	I_961297	SP:O96017	NM_007194.1	Hs.146329	BC004207.1	CHK2 checkpoint homolog (S. pombe)	417	418
CDC2	I_933293	SP:P06493	NM_001786.2; NM_033379.1	Hs.334562	BC014563.1	cell division cycle 2, G1 to S and G2 to M	419	420
CCNB1	I_958486	SP:P14635	NM_031966.1	Hs.23960	BC006510.1	cyclin B1	423	424
KNL4	I_959791	SP:Q14807	NM_007317.1	Hs.119324	AB017430.2	kinesin family member 22	425	426
LOC51053	I_966815	SP:O75496	NM_015895.1	Hs.234896	BC005389.1	geminin, DNA replication inhibitor	427	428
CKS2	I_931102	SP:P33552	NM_001827.1	Hs.83758	BC006458.1	CDC28 protein kinase regulatory subunit 2	429	430
PRC1	I_960183	GP:AAH03138.1	NM_003981.1	Hs.344037	BC003138.1	protein regulator of cytokinesis 1	431	432
TYMS	I_960396	SP:P04818	NM_001071.1	Hs.87491	X02308.1	thymidylate synthetase	433	434
E2-EPF	I_961201	SP:Q16763	NM_014501.1	Hs.396393	M91670.1	ubiquitin carrier protein	435	436
SMARCD1	I_931943	GP:AAH09368.1		Hs.79335	BC009368.1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 1	533	532
HSPC150	I_929756	GP:BAA91211.1	NM_014176.1	Hs.5199	AF160215.1	HSPC150 protein similar to ubiquitin-conjugating enzyme	437	438
LSM5	I_929834	SP:Q9Y4Y9	NM_012322.1	Hs.424908	AK024217.1	LSM5 homolog, U6 small nuclear RNA associated (S. cerevisiae)	439	440
SF3A3	I_1221922	SP:Q12874		Hs.77897	BC011523.1	splicing factor 3a, subunit 3, 60kDa	531	530
HEC	I_960629	GP:AAB80726.1	NM_006101.1	Hs.414407	AF017790.1	highly expressed in cancer, rich in leucine heptad repeats	441	442
FEN1	I_931399	SP:P39748	NM_004111.3	Hs.409065	BC000323.1	flap structure-specific endonuclease 1	443	444
ZWINT	I_933172	GP:AAC78629.1	NM_007057.2; NM_032997.1	Hs.42650	BC000411.1	ZW10 interactor	445	446
DTYMK	I_963220	GP:CAA38528.1	NM_012145.1	Hs.367821	X54729.1	deoxythymidylate kinase (thymidylate kinase)	449	450
RAD1	I_957256	GP:AAC35549.1	NM_002853.1	Hs.7179	AF030933.1	RAD1 homolog (S. pombe)	451	452

Gene Name	Agilent	Protein	RefSeq	Unigene	GenBank	Description	Amino Acid SEQ ID NO.	Nucleic Acid SEQ ID NO.
TACC3	I_957372	SP:Q9Y6A5	NM_006342.1	Hs.104019	AF093543.1	transforming, acidic coiled-coil containing protein 3	453	454
GTSE1	I_961917	GP:AAH06325.1	NM_016426.1	Hs.122552	BC006325.1	G-2 and S-phase expressed 1	455	456
RRM2	I_965619	SP:P31350	NM_001034.1	Hs.226390	X59618.1	ribonucleotide reductase M2		
RAB6KIFL	I_1110379	SP:Q95235	NM_005733.1	Hs.73625	AK025790.1	polypeptide	457	458
MCM7	I_929577	GP:BAA05839.1	NM_005916.1	Hs.438720	D28480.1	kinesin family member 20A	459	460
ANLN	I_929934	GP:AAF75796.1	NM_018685.1	Hs.62180	AF273437.1	MCM7 minichromosome maintenance deficient 7 (S. cerevisiae)	461	462
SNRPG	I_1100581	SP:Q15357	NM_003096.1	Hs.436656	BC022432.1	amlinin, actin binding protein (scraps homolog, Drosophila)	463	464
CSE1L	I_960930	SP:P55060	NM_001316.1	Hs.90073	AF053641.1	small nuclear ribonucleoprotein polypeptide G	465	466
TK1	I_960984	SP:P04183	NM_003258.1	Hs.164457	K02581.1	CSE1 chromosome segregation 1- like (yeast)	467	468
RRM1	I_930353	SP:P23921	NM_001033.1	Hs.383396	X59543.1	thymidine kinase 1, soluble	469	470
ASK	I_930306	GP:AAD45357.1	NM_006716.1	Hs.152759	AF160876.1	ribonucleotide reductase M1	471	472
SNRPA1	I_959930	SP:P09661	NM_003090.1	Hs.434901	BC022816.1	polypeptide activator of S phase kinase	473	474
TUBA4	I_933609	GP:BAB14767.1	NM_025019.1	Hs.287610	AK024002.1	small nuclear ribonucleoprotein polypeptide A'	475	476
BCL2A1	I_960129	SP:Q16548	NM_004049.2	Hs.227817	Y09397.1	tubulin, alpha 4	477	478
ERH	I_959366	SP:Q14259	NM_004450.1	Hs.433413	U66871.1	BCL2-related protein A1	479	480
TTK	I_966662	SP:P33981	NM_003318.1	Hs.169840	BC000633.1	enhancer of rudimentary homolog (Drosophila)	481	482
KNSL6	I_964064	SP:Q99661	NM_006845.2	Hs.69360	BC014924.1	TTK protein kinase	483	484
CDC45L	I_961013	SP:O75419	NM_003504.2	Hs.114311	AF081535.1	kinesin family member 2C	485	486
HIF4	I_957913	SP:P10412	NM_005321.1		M60748.1	CDC45 cell division cycle 45-like (S. cerevisiae)	487	488
RAD54B	I_1109914	GP:AAD34331.1	NM_012415.1	Hs.128501	AF112481.1	RAD54B homolog	489	490
HMG2	I_957632	SP:P26583	NM_002129.2	Hs.434953	X62534.1	high-mobility group box 2	491	492
CDC20	I_931677	GP:AAH09426.1		Hs.82906	BC000624.1	CDC20 cell division cycle 20 homolog (S. cerevisiae)	493	494
							535	534

Gene Name	Agilent	Protein	RefSeq	Unigene	GenBank	Description	Amino Acid SEQ ID NO.	Nucleic Acid SEQ ID NO.
MCM10	I_1100848	GP:CAB66774.1	NM_018518.1	Hs.198363	AL136840.1	MCM10 minichromosome maintenance deficient 10 (S. cerevisiae)	495	496
KPNB1	I_960232	SP:Q14974	NM_002265.1	Hs.439683	L38951.1	karyopherin (importin) beta 1	497	498
FOXMI	I_934617	GP:AACS1128.1	NM_021953.1	Hs.511941	U74612.1	forkhead box M1	499	500
USP15	I_932771	GP:BAA25455.2	NM_006313.1	Hs.339425	AF106069.1	ubiquitin specific protease 15	501	502
DHFR	I_1109838	SP:P00374	NM_000791.2	Hs.83765	BC000192.1	dihydrofolate reductase	503	504
DNMT1	I_961245	SP:P26358	NM_001379.1	Hs.202672	X63692.1	DNA (cytosine-5-)- methyltransferase 1	505	506

No cluster of genes upregulated were found in SLC/THAP1 expressing cells. In contrast, several clusters of genes downregulated by the SLC/THAP1 complex were discerned, which were not affected when the chemokine was expressed alone (Table 3A). Most of these genes were also downregulated by THAP1 without chemokine, however the chemokine greatly enhanced their down-regulation (co-repressor effect).

We identified ~ 120 candidate target genes (out of 17000 genes on the microarrays) that are downregulated in SLC/THAP1-overexpressing cells. One of these genes corresponds to THAP1 itself (FLJ10477), and many other genes correspond to predicted proteins with unknown functions. Strikingly, most of the genes encoding proteins with known functions (60 genes) that are downregulated by the SLC/THAP1 complex (Tables 3A and 4) correspond to genes encoding proteins previously linked to cell cycle/cell proliferation (Ishida et al. (2001) Mol Cell Biol 21:4684-4699; Whitfield et al. (2002) Mol Biol Cell 13:1977-2000; ): G2/M phase specific genes involved in mitosis (38 genes) and S phase specific genes involved in DNA replication or DNA repair (22 genes). Interestingly, many of these cell-cycle specific genes (26 genes, indicated in italics in Table 4) have previously shown to be regulated positively by the cell-cycle specific transcription factor E2F (Ishida et al. (2001) Mol Cell Biol 21:4684-4699; Ren et al. (2002) Genes Dev 16:245-256), suggesting that the SLC/THAP1 complex interfere some way with E2F-mediated activities. In addition to the cell cycle specific genes, genes encoding splicing factors (5 genes) and anti-apoptotic factors (2 genes including surviving) were also identified as target genes down-regulated by the SLC/THAP1 complex (Table 4). Together, these results indicated that the nuclear chemokine SLC/THAP1 complex modulate transcription profiles in human primary endothelial cells and appear to be a critical regulator of cell cycle/cell proliferation and/or survival.

**TABLE 4**  
**Target genes downregulated by the SLC/THAP1 complex**  
*(Genes indicated in italics are E2F target genes)*

G2 and M phase cluster Mitosis	S phase cluster DNA replication	Splicing	Apoptosis
<b>CKS1</b> (polypeptide SEQ ID NO: 367; nucleic acid SEQ ID NO: 368)	<b>H1F5 (H1 histone family)</b> (polypeptide SEQ ID NO: 381; nucleic acid SEQ ID NO: 382)	<b>U1snRNPBP</b> Gi/5902144 (polypeptide SEQ ID NO: 377; nucleic acid SEQ ID NO: 378)	<b>BIRC5/survivin</b> (polypeptide SEQ ID NO: 395; nucleic acid SEQ ID NO: 396)
<b>PTTG2/securin2</b> (polypeptide SEQ ID NO: 369 nucleic acid SEQ ID NO: 370)	<b>H1F3 (H1 histone family)</b> (polypeptide SEQ ID NO: 389; nucleic acid SEQ ID NO: 390)	gi/13027642 (transcript variant 2 polypeptide SEQ ID NO: 379; transcript variant 2 nucleic acid SEQ ID NO: 380)	<b>Bcl2-related protein</b> (polypeptide SEQ ID NO: 479; nucleic acid SEQ ID NO: 480)
<b>CDKN3</b> (polypeptide SEQ ID NO: 371; nucleic acid SEQ ID NO: 372)	<b>CDCA7 (c-myc target JPO1)</b> (polypeptide SEQ ID NO: 393; nucleic acid SEQ ID NO: 394)		
<b>BUB1</b> (polypeptide SEQ ID NO: 373; nucleic acid SEQ ID NO: 374)	<b>DUT (dUTP phosphatase)</b> (polypeptide SEQ ID NO: 409; nucleic acid SEQ ID NO: 410)	<b>LSM5 (U6 snRNA associated Sm-like protein)</b> (polypeptide SEQ ID NO: 439; nucleic acid SEQ ID NO: 440)	
<b>HMMR/RHAMM</b> (polypeptide SEQ ID NO: 375; nucleic acid SEQ ID NO: 376)	<b>LOC51053/Geminin</b> (polypeptide SEQ ID NO: 427; nucleic acid SEQ ID NO: 428)	<b>SF3A3 (spliceosome-associated protein 61)</b> (polypeptide SEQ ID NO: 531; nucleic acid SEQ ID NO: 530)	
<b>PTTG3/securin3</b> (polypeptide SEQ ID NO: 383; nucleic acid SEQ ID NO: 384)	<b>TYMS (thymidylate synthetase)</b> (polypeptide SEQ ID NO: 433; nucleic acid SEQ ID NO: 434)	<b>SnRNPG (Sm core protein G)</b> (polypeptide SEQ ID NO: 465; nucleic acid SEQ ID NO: 466)	
<b>TOPK (PDZ-binding kinase)</b>	<b>HSPC150/Ubc13p</b>	<b>SnRNPA1</b>	

G2 and M phase cluster Mitosis	S phase cluster DNA replication	Splicing	Apoptosis
(polypeptide SEQ ID NO: 385; nucleic acid SEQ ID NO: 386) <i>PTTG1/securin1</i> (polypeptide SEQ ID NO: 387; nucleic acid SEQ ID NO: 388) <i>KIAA0008/HURP</i> (polypeptide SEQ ID NO: 391; nucleic acid SEQ ID NO: 392) <i>BIRC5/survivin</i> (polypeptide SEQ ID NO: 395; nucleic acid SEQ ID NO: 396) <i>CNAP1</i> (polypeptide SEQ ID NO: 397; nucleic acid SEQ ID NO: 398) <i>USP16</i> (polypeptide SEQ ID NO: 401; nucleic acid SEQ ID NO: 402) <i>MAD2L1</i> (polypeptide SEQ ID NO: 403; nucleic acid SEQ ID NO: 404)	(polypeptide SEQ ID NO: 437; nucleic acid SEQ ID NO: 438) <i>FEN1 (flap endonuclease 1)</i> (polypeptide SEQ ID NO: 443; nucleic acid SEQ ID NO: 444) <i>DTYMK (dTMP kinase)</i> (polypeptide SEQ ID NO: 449; nucleic acid SEQ ID NO: 450) <i>RAD1</i> (polypeptide SEQ ID NO: 451; nucleic acid SEQ ID NO: 452) <i>RRM2 (ribonucleotide reductase 2)</i> (polypeptide SEQ ID NO: 457; nucleic acid SEQ ID NO: 458) <i>MCM7</i> (polypeptide SEQ ID NO: 461; nucleic acid SEQ ID NO: 462) <i>TK1 (thymidine kinase 1)</i> (polypeptide SEQ ID NO: 469; nucleic acid SEQ ID NO: 470)  <i>RRM1 (ribonucleotide reductase 1)</i> (polypeptide SEQ ID NO: 471;	(polypeptide SEQ ID NO: 475; nucleic acid SEQ ID NO: 476)	
<i>CDCA1</i> (polypeptide SEQ ID NO: 405;			

G2 and M phase cluster Mitosis	S phase cluster DNA replication	Splicing	Apoptosis
nucleic acid SEQ ID NO: 406) <b>BUB1B/BUBR1</b> (polypeptide SEQ ID NO: 407; nucleic acid SEQ ID NO: 408) <b>KNSL7</b> (kinesin-like 7) (polypeptide SEQ ID NO: 411; nucleic acid SEQ ID NO: 412) <b>KNSL1</b> (kinesin-like 1) (polypeptide SEQ ID NO: 413; nucleic acid SEQ ID NO: 414) <b>CCNB2</b> ( <i>cyclin B2</i> ) (polypeptide SEQ ID NO: 415; nucleic acid SEQ ID NO: 416) <b>CHEK2</b> (CHK2 checkpoint) (polypeptide SEQ ID NO: 417; nucleic acid SEQ ID NO: 418) <b>CDC2</b> (Isoform 1: polypeptide SEQ ID NO: 419; variant 2: nucleic acid SEQ ID NO: 420) isoform 2: (polypeptide SEQ ID NO: 421; variant 2: nucleic acid SEQ ID NO:	nucleic acid SEQ ID NO: 472) <i>ASK</i> ( <i>activator of S phase kinase</i> ) (polypeptide SEQ ID NO: 473; nucleic acid SEQ ID NO: 474) <b>ERH</b> (polypeptide SEQ ID NO: 481; nucleic acid SEQ ID NO: 482) <b>CDC45L</b> (polypeptide SEQ ID NO: 487; nucleic acid SEQ ID NO: 488) <b>H1F4</b> (H1 histone family) (polypeptide SEQ ID NO: 489; nucleic acid SEQ ID NO: 490) <b>RAD54B</b> (polypeptide SEQ ID NO: 491; nucleic acid SEQ ID NO: 492) <b>MCM10</b> (polypeptide SEQ ID NO: 495; nucleic acid SEQ ID NO: 496) <b>I_1109838/DHFR</b> (polypeptide SEQ ID NO: 503; nucleic acid SEQ ID NO: 504)		

G2 and M phase cluster Mitosis	S phase cluster DNA replication	Splicing	Apoptosis
<b>422)</b> <b>CCNB1</b> ( <i>cyclin B1</i> ) (polypeptide SEQ ID NO: 423; nucleic acid SEQ ID NO: 424) <b>KNSL4</b> ( <i>kinesin-like 4</i> ) (polypeptide SEQ ID NO: 425; nucleic acid SEQ ID NO: 426) <b>CKS2</b> (polypeptide SEQ ID NO: 429; nucleic acid SEQ ID NO: 430) <b>PRCI</b> (polypeptide SEQ ID NO: 431; nucleic acid SEQ ID NO: 432) <b>E2-EPF</b> (polypeptide SEQ ID NO: 435; nucleic acid SEQ ID NO: 436) <b>SMARCD1</b> (polypeptide SEQ ID NO: 533; nucleic acid SEQ ID NO: 532)  <b>HEC</b> (polypeptide SEQ ID NO: 441; nucleic acid SEQ ID NO: 442) <b>ZWINT</b>	<b>DNMT1</b> (DNA methyltransferase) (polypeptide SEQ ID NO: 505; nucleic acid SEQ ID NO: 506)		



G2 and M phase cluster Mitosis	S phase cluster DNA replication	Splicing	Apoptosis
<p>(polypeptide SEQ ID NO: 445; nucleic acid SEQ ID NO: 446) (polypeptide SEQ ID NO: 447; variant 2 nucleic acid SEQ ID NO: 448)</p> <p><b>TACC3</b></p> <p>(polypeptide SEQ ID NO: 453; nucleic acid SEQ ID NO: 454)</p> <p><b>GTSE1</b></p> <p>(polypeptide SEQ ID NO: 455; nucleic acid SEQ ID NO: 456)</p> <p><b>RAB6KIFL (raklinsin 6)</b></p> <p>(polypeptide SEQ ID NO: 459; nucleic acid SEQ ID NO: 460)</p> <p><b>ANLN (anilin)</b></p> <p>(polypeptide SEQ ID NO: 463; nucleic acid SEQ ID NO: 464)</p> <p><b>CSE1L (importin alpha)</b></p> <p>(polypeptide SEQ ID NO: 467; nucleic acid SEQ ID NO: 468)</p> <p><b>TUBA4</b></p> <p>(polypeptide SEQ ID NO: 477; nucleic acid SEQ ID NO: 478)</p> <p><b>TTK (dual specificity kinase)</b></p> <p>(polypeptide SEQ ID NO: 483;</p>			

G2 and M phase cluster Mitosis	S phase cluster DNA replication	Splicing	Apoptosis
nucleic acid SEQ ID NO: 484) <b>KNSL6</b> (kinesin-like 6) (polypeptide SEQ ID NO: 485; nucleic acid SEQ ID NO: 486) <b>HMG2</b> (polypeptide SEQ ID NO: 493; nucleic acid SEQ ID NO: 494) <b>CDC20</b> ( <i>p55CDC</i> ) (polypeptide SEQ ID NO: 535; nucleic acid SEQ ID NO: 534)			

## EXAMPLE 48

Chemokines SLC/CCL21 and MIG/CXCL9 modulate  
transcription of pro-inflammatory chemokine genes

5 To examine the expression of nuclear chemokines SLC/CCL21 and MIG/CXCL9, we performed DNA microarrays analysis of HUVEC cells transduced with SLC/CCL21 or MIG/CXCL9 retrovirus vectors or MCS control vector. Cluster analysis was performed based on similarity of expression patterns of genes.

10 Table 5A lists 5 genes encoding pro-inflammatory chemokines that are downregulated by chemokines SLC/CCL21 and MIG/CXCL9 in human primary endothelial cells, by the above-describe microarray experiments. For each chemokine gene, the fold changes, p values and signal intensities obtained in the two microarray experiments are indicated.

Table 5B lists database accession numbers and SEQ ID NOs. for each gene and corresponding polypeptide listed in Table 3A.

15

TABLE 5A

Sequence Name(s)	Experiment Name Sequence Description	SLC/MCSvsMCS			MIG/MCSvsMC		
		Fold Change	P-Value	Intensity 1	Fold Change	P-Value	Intensity 1
GRO1	Growth related oncogene (melanoma growth stimulating activity), a CXC chemokine that binds interleukin 8 receptor to mobilize intracellular calcium, acts as a leukocyte mitogenic factor with growth-regulatory and chemotactic properties during inflammation	-1.86	3.73E-07	41433.3	-2.47	1.92E-30	41799.53
GRO2	Macrophage inflammatory protein 2, a member of the C-X-C chemokine family, acts as a neutrophil chemoattractant and epithelial cell mitogen	-1.78	1.36E-13	10921.9	-2.22	1.00E-36	10778.54
IL8	Interleukin 8, a cytokine that plays a role in chemoattraction and activation of neutrophils, involved in immune and inflammatory responses	-1.74	1.77E-16	33425.4	-1.84	8.38E-24	32784.91
GRO3	Melanoma growth stimulating activity gamma, a chemokine and mitogenic factor, activates neutrophils and induces chemotaxis, may be involved in the inflammatory response	-1.49	0.04	948.15	-1.82	6.23E-07	874.77
SCYA2	Cytokine A2, CC chemokine that attracts monocytes, memory T-cells, natural killer cells and endothelial cells, plays a role in the inflammatory response to infection and in inflammatory diseases including arthritis, multiple sclerosis and atherosclerosis	-1.24	0.09	9503.27	-1.69	1.51E-05	10442.52
							6173.76

TABLE 5B

Gene Name	Agilent	Protein	RefSeq	GenBank	Amino Acid SEQ ID NO.	Nucleic Acid SEQ ID NO.
GRO1	I_957623	SP:P09341	NM_001511.1	BC011976.1	507	508
GRO2	I_957614	SP:P19875	NM_002089.1	BC015753.1	509	510
IL8	I_957620	SP:P10145	NM_000584.1	Y00787.1	511	512
GRO3	I_957616	SP:P19876	NM_002090.1	BC016308.1	513	514
SCYA2	I_959180	SP:P13500	NM_002982.1	M24545.1	515	516

The chemokines SLC/CCL21 or MIG/CXCL9 expressed alone, induced changes in HUVEC gene expression profile characterized by distinct clusters of genes upregulated or downregulated. Interestingly, the main cluster of genes down-regulated by both SLC/CCL21 or MIG/CXCL9 corresponded to genes encoding pro-inflammatory chemokines GRO1/CXCL1, GRO2/CXCL2, GRO3/CXCL3, IL8/CXCL8 and MCP1/CCL2 (Table 5A). Together, these results indicated that nuclear chemokines SLC/CCL21 and MIG/CXCL9 are able to modulate transcription profiles in human primary endothelial cells and may have anti-inflammatory effects by inhibiting expression of pro-inflammatory chemokines.

## EXAMPLE 49

10 Construction of Adenovirus Vectors for Expressing THAP-Family Polypeptides and Chemokines

This example illustrates the construction of adenovirus vectors comprising nucleic acids encoding THAP1, SLC and MIG. It will be appreciated that these methods can be applied to other THAP-family polypeptides, chemokines and/or chemokine receptors as desired.

The full-length cDNA encoding human THAP1 (SEQ ID NO: 160) is amplified from human cDNA. Similarly, mature forms (forms lacking a signal peptide) of the chemokines SLC and MIG can be amplified from human cDNA. The resulting PCR products are purified from an agarose gel and then ligated into a TA-cloning vector, such as pCR2.1 (Invitrogen, Carlsbad, Calif.). Once the cDNA insert sequence is verified by sequence analysis, the plasmid containing the insert of interest is digested to remove the cDNA insert, which is then blunt-ended with T4 DNA polymerase, gel purified and ligated into the EcoRV site of the adenoviral shuttle vectors pAvS6a to form pAvS6a-THAP1, pAvS6a-SLC or pAcS6a-MIG. Finally, a fragment which contains the cDNA insert of interest is removed from each of the pAvS6a recombinant vectors using appropriate restriction enzymes and then subcloned into pAvS6alx (a shuttle vector containing lox site, Genetic Therapy, Inc., Gaithersburg, Md.) to generate, for example, pAvhTHAP1Ix. The expression cassettes thereby generated include the gene of interest, a constitutive RSV promoter, a 198 bp fragment containing the adeno-tripartite leader sequence, lox recombination sequence, and an SV40 early polyadenylation signal.

The recombinant adenovirus encoding human THAP1 (Av3hTHAP1), SLC (Av3hSLC) or MIG (Av3hMIG) are constructed by a rapid vector generation protocol using Cre recombinase-mediated recombination of two lox-site containing plasmids, pSQ3 (containing the right hand portion of the adenoviral vector genome), and the adenoviral shuttle plasmid pAvhTHAP1Ix (containing the left end of the viral genome and the hTHAP1 expression cassette) pAvhSLC1x or pAvhMIG1x. The pSQ3 (digested with ClaI), pAvhTHAP1Ix, pAvhSLC1x or pAvhMIG1x (linearized with NotI), and the Cre-encoding plasmid, pC-Cre3.1, are cotransfected using CaPO<sub>4</sub> (Promega's Profection kit) into S8 cells (A549 cells stably transfected with E1/E2a regions under dexamethasone inducible promoters (Gorziglia et al., J. Virol. 6:41734178, 1996). Following treatment with dexamethasone the plasmids are joined by Cre-mediated recombination, generating

the adenovirus encoding THAP1 (Av3hTHAP1), SLC (Ac3hSLC) or MIG (Av3hMIG). A control vector, Av3Null is generated in a similar manner, but lacks a transgene.

To amplify the virus, the S8 cells are harvested a week after transfection and passaged until a cytopathic effect (CPE) is observed. For the passage, cells are freeze/thawed to obtain a crude viral lysate (CVL), which is centrifuged to remove the cell debris and then used to infect fresh S8 cells. Cells are harvested when CPE is observed (typically after one week). DNA is isolated from the CVL and the appropriate cre-lox mediated recombination event is confirmed by restriction digest. For purification of the vector, cell pellets are freeze/thawed and the cell debris are pelleted by centrifugation. The supernatant is loaded on a discontinuous Cesium Chloride gradient (1.25 g/ml CsCl and 1.4 g/ml CsCl) and centrifuged for 1 hr at 28,000 rpm (in a SW28 swing bucket rotor). The bottom viral band is pulled from the gradient and centrifuged on a CsCl continuous gradient (1.33 g/ml CsCl) overnight at 60,000 rpm (in an NVT-65 rotor). The purified viral band is pulled from the gradient, glycerol is added to a final concentration of 10% and the mixture is then dialyzed in 200 mM Tris pH 8.0, 50 mM Hepes, 10% glycerol. The concentration of vector can be determined by spectrophotometric analysis (Mittereder et al., J. Virol. 70:7498-7509, 1996). Purified vector is then aliquoted and stored at -70°C.

Av3hTHAP1, AV3hSLC and Av3hMIG vector expression is examined in HUVEC cells. The cells are treated for 1 hour with varying multiplicities of infection of Av3hTHAP1, AV3hSLC, Av3hMIG or Av3Null or left untreated. Two days following treatment, cell extracts are prepared and Western blot analysis is performed using an antibody specific for THAP1, SLC or MIG. The biological activity of the expressed THAP1 protein is confirmed using the serum starvation assays as described in Examples 10 and 11. Alternatively, the effect of THAP1, SLC, MIG, or combinations of these polypeptides on gene transcription can be determined by comparing transcriptional activities of cells transfected with one or more of Av3hTHAP1, AV3hSLC, Av3hMIG with the transcriptional activities of cells transfected with Av3Null. Assays for determining gene expression as well as several genes modified by THAP1 and THAP1/chemokine complexes have been described in Examples 44-47.

It will be appreciated by one of ordinary skill in the art, that vectors which express a both a chemokine as well as a THAP-family polypeptide or biologically active fragment thereof can also be constructed using the methods described above. Additionally, a skilled artisan will recognize that vectors other than adenovirus vectors can be use generate constructs capable of expressing a chemokine and/or a THAP-family polypeptide or a biologically active fragment thereof. Such vectors include, but are not limited to, adenovirus associated vectors, lentivirus vectors and retrovirus vectors. Additionally, non-viral vector may be used.

## EXAMPLE 50

Expression of THAP-Family Polypeptides and  
Chemokines in a Mouse Model of Rheumatoid Arthritis

This example illustrates the use of adenovirus vectors to deliver nucleic acids encoding THAP1, SLC, MIG or combinations of these polypeptides to inflamed tissue in a mouse model for rheumatoid arthritis, the well-known collagen-induced arthritis model.

Male DBA/1 mice are prepared as in Example 36 above. For viral dosing of mice, the DBA/1 mice are administered recombinant adenoviruses via tail vein injection using a 0.5 ml tuberculin syringe at doses of  $0.6-1.2 \times 10^{11}$  viral particles/animal. Four groups of animals (n=5-15/group) are treated with either Av3hTHAP1, Av3hSLC, Av3hMIG, combinations of these recombinant viruses, Av3Null or buffer only.

The incidence and severity of arthritis is monitored in a blind study. Each paw is assigned a score from 0 to 4 as follows: 0=normal; 1=swelling in 1 to 3 digits; 2=mild swelling in ankles, forepaws, or more than 3 digits; 3=moderate swelling in multiple joints; 4=severe swelling with loss of function. Each paw is totaled for a cumulative score/mouse. The cumulative scores are then totaled for mice in each group for a mean clinical score. The capacity for THAP1 or THAP1/chemokine combinations to reduce the disease incidence and severity of arthritis is determined by comparison of the treatment groups to the control groups.

It will be appreciated by one of ordinary skill in the art that expression of a chemokine and/or a THAP-family polypeptide or a biologically active fragment thereof can be used to ameliorate the symptoms associated with any THAP-related condition. In some embodiments such expression can be the result of gene therapy.

The methods, compositions, and devices described herein are presently representative of preferred embodiments and are exemplary and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention and are defined by the scope of the disclosure. Accordingly, it will be apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention.

As used in the claims below and throughout this disclosure, by the phrase "consisting essentially of" is meant including any elements listed after the phrase, and limited to other elements that do not interfere with or contribute to the activity or action specified in the disclosure for the listed elements. Thus, the phrase "consisting essentially of" indicates that the listed elements are required or mandatory, but that other elements are optional and may or may not be present depending upon whether or not they affect the activity or action of the listed elements.

REFERENCES

Numerous literature and patent references have been cited in the present application.

For some references, the complete citation is in the body of the text. For other references the citation in the body of the text is by author and year, the complete citation being as follows:

- 5
- Alcalay, M., Tomassoni, L., Colombo, E., Stoldt, S., Grignani, F., Fagioli, M., Szekely, L., Helin, K., and Pelicci, P. G. (1998). The promyelocytic leukemia gene product (PML) forms stable complexes with the retinoblastoma protein. *Mol Cell Biol* 18, 1084-93.
- Arcone, et al., *Nucl. Acids Res.*, 16(8): 3195-3207, 1988.
- 10 • Baichwal and Sugden, In: *Gene Transfer*, Kucherlapati R, ed., New York, Plenum Press, pp. 117-148, 1986.
- Barradas, M., Monjas, A., Diaz-Meco, M. T., Serrano, M., and Moscat, J. (1999). The downregulation of the pro-apoptotic protein Par4 is critical for Ras-induced survival and tumor progression. *Embo J* 18, 6362-9.
- 15 • Bartlett et al., *Proc. Nat Acad Sci. USA*, 93:8852-8857, 1996.
- Benvenisty and Neshif, *Proc. Nat Acad Sci. USA*, 83:9551-9555, 1986.
- Berra, E., Municio, M. M., Sanz, L., Frutos, S., Diaz-Meco, M. T., and Moscat, J. (1997). Positioning atypical protein kinase C isoforms in the UV-induced apoptotic signaling cascade. *Mol Cell Biol* 17, 4346-54.
- 20 • Bloch, D. B., Chiche, J. D., Orth, D., de la Monte, S. M., Rosenzweig, A., and Bloch, K. D. (1999). Structural and functional heterogeneity of nuclear bodies. *Mol Cell Biol* 19, 4423-30.
- Boghaert E.R. et al, *Cell Growth Differ.* 1997 Aug;8(8):881-90
- Burge, C., Karlin, S. (1997) Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268, 78-94
- 25 • Carter and Flotte, *Ann. N. Y Acad Sci.*, 770:79-90, 1995.
- Chattedee, et al., *Ann. N. Y Acad Sci.*, 770:79-90, 1995.
- Chen and Okayama, *Mol. Cell Biol.*, 7:2745-2752, 1987.
- Coffin, In: *Virology*, ed., New York: Raven Press, pp. 143 7-15 00, 1990.
- Coupar et al., *Gene*, 68:1-10, 1988.
- 30 • Dani, et al., *J Biol. Chem.*, 264:10119-10125, 1989.
- Diaz-Meco, M. T., Municio, M. M., Frutos, S., Sanchez, P., Lozano, J., Sanz, L., and Moscat, J. (1996). The product of Par4, a gene induced during apoptosis, interacts selectively with the atypical isoforms of protein kinase C. *Cell* 86, 777-86.
- Dubensky et al., *Proc. Nat'l Acad Sci. USA*, 81:7529-7533, 1984.
- 35 • Fechheimer et al., *Proc. Nat Acad. Sci. USA*, 84:8463-8467, 1987.
- Felsenstein, J. 1989. PHYLIP -- Phylogeny Inference Package (Version 3.2). *Cladistics* 5: 164-166.



- Ferkol et al., *FASEB J.*, 7:1081-1091, 1993.
- Ferrari et al., *J Virol.*, 70:3227-3234, 1996.
- Fisher et al., *J Virol.*, 70:520-532, 1996.
- Flotte et al., *Proc. Nat'l Acad. Sci. USA*, 90:10613-10617, 1993.
- 5 • Fogal, V., Gostissa, M., Sandy, P., Zacchi, P., Sternsdorf, T., Jensen, K., Pandolfi, P. P., Will, H., Schneider, C., and Del Sal, G. (2000). Regulation of p53 activity in nuclear bodies by a specific PML isoform. *Embo J* 19, 6185-95.
- Fraley et al., *Proc. Nat'l Acad. Sci. USA*, 76:3348-3352, 1979.
- Friedmann, "Progress toward human gene therapy," *Science*, 244:1275-1281, 1989.
- 10 • Ghosh and Bachhawat, In: *Liver diseases, targeted diagnosis and therapy using specific receptors and ligands*, (Wu G, Wu C ed.), New York: Marcel Dekker, pp. 87 104, 1991.
- Girard, J. P., Baekkevold, E. S., Feliu, J., P, B., and Amalric, F. (1999). Molecular cloning and functional analysis of SUT-1, a sulfate transporter from human high endothelial venules. *Proc Natl Acad Sci USA* 96:12772-12777.
- 15 • Girard, J. P., Baekkevold, E. S., Yamanaka, T., Haraldsen, G., Brandtzaeg, P., and Amalric, F. (1999). Heterogeneity of endothelial cells : the specialized phenotype of human high endothelial venules characterized by suppression subtractive hybridization. *Am J Pathol* 155:2043-2055.
- Girard, J. P., and Springer, T. A. (1995). Cloning from purified high endothelial venule cells of hevin, a close relative of the antiadhesive extracellular matrix protein SPARC. *Immunity* 2, 113-123.
- 20 • Goodman et al., *Blood*, 84:1492-1500, 1994.
- Gopal, *Mol. Cell Biol.*, 5:1188-1190, 1985.
- Gossen and Bujard, *Proc. Nat Acad. Sci. USA*, 89:5547-5551, 1992.
- 25 • Gossen et al., *Science*, 268:1766-1769, 1995.
- Graham and Van Der Eb, *Virology*, 52:456-467, 1973.
- Guo, Q., Fu, W., Xie, J., Luo, H., Sells, S. F., Geddes, J. W., Bondada, V., Rangnekar, V. M., and Mattson, M. P. (1998). Par4 is a mediator of neuronal degeneration associated with the pathogenesis of Alzheimer disease. *Nat Med* 4, 957-62.
- 30 • Hay et al., *J Mol. Biol.*, 175:493 -510, 1984.
- Hearing and Shenk, *J Mol. Biol.* 167:809-822, 1983.
- Hearing et al., *J Virol.*, 67:2555-2558, 1987.
- Higgins, D. G., Thompson, J. D. and Gibson, T. J. (1996) Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.*, 266, 383-402.
- 35 • Horowitz, Yandell, Park, Canning, Whyte, Buchkovich, Harlow, Weinberg, Dryja, "Point mutational inactivation of the retinoblastoma antioncogene," *Science* 243, 937 940, 1989.

- Hunt et al., Proc. Nat Acad. Sci. USA, 83:3786-3790, 1986.
- Ishov, A. M., Sotnikov, A. G., Negorev, D., Vladimirova, O. V., Neff, N., Kamitani, T., Yeh, E. T., Strauss, J. F., 3rd, and Maul, G. G. (1999). PML is critical for ND10 formation and recruits the PML-interacting protein daxx to this nuclear structure when modified by SUMO-1. *J Cell Biol* 147, 221-34.
- Jareborg N, Birney E, Durbin R. *Genome Res.* 1999 Sep;9(9):815-24.
- Jentsch, S., and Pyrowolakis, G. (2000). Ubiquitin and its kin: how close are the family ties? *Trends Cell Biol* 10, 335-42.
- Johnstone, R. W., See, R. H., Sells, S. F., Wang, J., Muthukumar, S., Englert, C., Haber, D. A., Licht, J. D., Sugrue, S. P., Roberts, T., Rangnekar, V. M., and Shi, Y. (1996). A novel repressor, Par4, modulates transcription and growth suppression functions of the Wilms' tumor suppressor WT1. *Mol Cell Biol* 16, 6945-56.
- Joki et al., *Human Gene Ther.*, 6:1507-1513, 1995.
- Kageyama, et al., *J Biol. Chem.*, 262(5):2345-2351, 1987.
- Kaneda et al., *Science*, 243:375-378, 1989.
- Kaplitt et al., *Nat Genet.*, 8:148-153, 1994.
- Kato et al., *J Biol. Chem.*, 266:3361-3364, 1991.
- Kessler et al., *Proc. Nat Acad. Sci. USA*, 93:14082-14087, 1996.
- Klein et al., *Nature*, 327:70-73, 1987.
- Koeberl et al., *Proc. Nat 7 Acad. Sci. USA*, 94:1426-1431, 1997.
- Korhonen et al., *Blood*, 86:1828-1835, 1995.
- Lallemand-Breitenbach, V., Zhu, J., Puvion, F., Koken, M., Honore, N., Doubeikovsky, A., Duprez, E., Pandolfi, P. P., Puvion, E., Freemont, P., and de The, H. (2001). Role of Promyelocytic Leukemia (PML) Sumolation in Nuclear Body Formation, 11S Proteasome Recruitment, and As(2)O(3)-induced PML or PML/Retinoic Acid Receptor alpha Degradation. *J Exp Med* 193, 1361-72.
- LaMorte, V. J., Dyck, J. A., Ochs, R. L., and Evans, R. M. (1998). Localization of nascent RNA and CREB binding protein with the PML-containing nuclear body. *Proc Natl Acad Sci U S A* 95, 4991-6.
- Levrero et al., *Gene*, 10 1: 195-202, 199 1.
- Li, H., Leo, C., Zhu, J., Wu, X., O'Neil, J., Park, E. J., and Chen, J. D. (2000). Sequestration and inhibition of Daxx-mediated transcriptional repression by PML. *Mol Cell Biol* 20, 1784-96.
- Mann et al., *Cell*, 33:153-159, 1983.
- Mattson, M. P., Duan, W., Chan, S. L., and Camandola, S. (1999). Par4: an emerging pivotal player in neuronal apoptosis and neurodegenerative disorders. *J Mol Neurosci* 13, 17-30.
- Maul, G. G., Negorev, D., Bell, P., and Ishov, A. M. (2000). Review: properties and assembly mechanisms of ND10, PML bodies, or PODs. *J Struct Biol* 129, 278-87.

- Melchior, F. (2000). SUMO--nonclassical ubiquitin. *Annu Rev Cell Dev Biol* 16, 591-626.
- Mizukami et al., *Virology*, 217:124-130, 1996.
- Muller, S., Matunis, M. J., and Dejean, A. (1998). Conjugation with the ubiquitin-related modifier SUMO-1 regulates the partitioning of PML within the nucleus. *Embo J* 17, 61-70.
- 5 • McCown et al., *Brain Res.*, 713:99-107, 1996.
- Nicolas and Rubenstein, In: *Vectors: A survey of molecular cloning vectors and their uses*, Rodriguez and Denhardt (eds.), Stoneham: Butterworth, pp. 493-513, 1988.
- Nicolau and Sene, *Biochim. Biophys. Acta*, 721:185-190, 1982.
- Nicolau et al., *Methods Enzymol.*, 149:157-176, 1987.
- 10 • Oliviero, et al., *EMBO J*, 6(7):1905-1912, 1987.
- Page, G., Kogel, D., Rangnekar, V., and Scheidtmann, K. H. (1999). Interaction partners of Dlk/ZIP kinase: co-expression of Dlk/ZIP kinase and Par4 results in cytoplasmic retention and apoptosis. *Oncogene* 18, 7265-73.
- Pape and Kim, *Mol. Cell. Biol.*, 974-982, 1989.
- 15 • Paskind et al., *Virology*, 67:242-248, 1975.
- Perales et al., *Proc. Nat Acad Sci.* 91:4086-4090, 1994.
- Ping et al., *Microcirculation*, 3:225-228, 1996.
- Poli and Cortese, *Proc. Nat'l Acad Sci. USA*, 86:8202-8206, 1989.
- Potter et al., *Proc. Nat Acad. Sci. USA*, 81:7161-7165, 1984.
- 20 • Prowse and Baumann, *Mol Cell Biol*, 8(1):42-51, 1988.
- Quignon, F., De Bels, F., Koken, M., Feunteun, J., Ameisen, J. C., and de The, H. (1998). PML induces a novel caspase-independent death process. *Nat Genet* 20, 259-65.
- Radler et al., *Science*, 275:810-814, 1997. Rb Protein Inhibits p84N5-induced Apoptosis, 10:3261, October 1999.
- 25 • Ridgeway, In: *Vectors: A survey of molecular cloning vectors and their uses*, Rodriguez RL, Denhardt DT, ed., Stoneham: Butterworth, pp. 467-492, 1988.
- Rippe et al., *Mol. Cell Biol.*, 10:689-695, 1990.
- Ron et al., *Mol. Cell. Biol.*, 2887-2895, 1991.
- Roux et al., *Proc. Nat Acad. Sci. USA*, 86:9079-9083, 1989.
- 30 • Ruggero, D., Wang, Z. G., and Pandolfi, P. P. (2000). The puzzling multiple lives of PML and its role in the genesis of cancer. *Bioessays* 22, 827-35.
- Sells, S. F., Han, S. S., Muthukkumar, S., Maddiwar, N., Johnstone, R., Boghaert, E., Gillis, D., Liu, G., Nair, P., Monnig, S., Collini, P., Mattson, M. P., Sukhatme, V. P., Zimmer, S. G., Wood, D. P., Jr., McRoberts, J. W., Shi, Y., and Rangnekar, V. M. (1997). Expression and function of the leucine zipper protein Par4 in apoptosis. *Mol Cell Biol* 17, 3823-32.
- 35 • Spiegelman, et al., *J Biol. Chem.*, 264(3), 1811-1815, 1989.

- Sternsdorf, T., Jensen, K., Reich, B., and Will, H. (1999). The nuclear dot protein sp100, characterization of domains necessary for dimerization, subcellular localization, and modification by small ubiquitin-like modifiers. *J Biol Chem* 274, 12555-66.
- Tartaglia, Ayres, Wong, and Goeddel, "A novel domain within the 55 kd TNF receptor signals cell death," *Cell* 74, 845-853, 1993.
- Temin, In: *Gene Transfer*, Kucherlapati (ed.), New York: Plenum Press, pp. 149-188, 1986.
- Tibbetts *Cell*, 12:243-249, 1977.
- Tur-Kaspa et al., *Mol. Cell Biol.*, 6:716-718, 1986.
- Wagner et al., *Proc. Nat'l Acad. Sci.* 87(9):3410-3414, 1990.
- Walther and Stein, *J Mol. Med.*, 74:3 79-3 92, 1996.
- Wang, Z. G., Delva, L., Gaboli, M., Rivi, R., Giorgio, M., Cordon-Cardo, C., Grosveld, F., and Pandolfi, P. P. (1998). Role of PML in cell growth and the retinoic acid pathway. *Science* 279, 1547-51.
- Wang, Z. G., Ruggero, D., Ronchetti, S., Zhong, S., Gaboli, M., Rivi, R., and Pandolfi, P. P. (1998). PML is essential for multiple apoptotic pathways. *Nat Genet* 20, 266-72.
- Watt et al., *Proc. Nat'l Acad. Sci.*, 83(2): 3166-3170, 1986.
- Wilson, et al., *Mol. Cell. Biol.*, 6181-6191, 1990.
- Wong et al., *Gene*, 10:87-94, 1980.
- Wu and Wu, *Adv. Drug Delivery Rev.*, 12:159-167, 1993.
- Wu and Wu, *Biochem.*, 27:887-892, 1988.
- Wu and Wu, *J Biol. Chem.*, 262:4429-4432, 1987.
- Xiao et al., *J Virol.*, 70:8098-8108, 1996.
- Yang et al., *Proc. Nat Acad. Sci USA*, 87:9568-9572, 1990.
- Yeh, E. T., Gong, L., and Kamitani, T. (2000). Ubiquitin-like proteins: new wines in new bottles. *Gene* 248, 1-14.
- Zechner, et al., *Mol. Cell. Biol.*, 23 94-2401, 198 8.
- Zhong, S., Hu, P., Ye, T. Z., Stan, R., Ellis, N. A., and Pandolfi, P. P. (1999). A role for PML and the nuclear body in genomic stability. *Oncogene* 18, 7941-7.
- Zhong, S., Muller, S., Ronchetti, S., Freemont, P. S., Dejean, A., and Pandolfi, P. P. (2000). Role of SUMO-1-modified PML in nuclear body formation. *Blood* 95, 2748-52.
- Zhong, S., Salomoni, P., and Pandolfi, P. P. (2000). The transcriptional role of PML and the nuclear body. *Nat Cell Biol* 2, E85-90.
- Zhong, S., Salomoni, P., Ronchetti, S., Guo, A., Ruggero, D., and Pandolfi, P. P. (2000). Promyelocytic leukemia protein (PML) and Daxx participate in a novel nuclear pathway for apoptosis. *J Exp Med* 191, 631-40.

WHAT IS CLAIMED IS:

1. A method of modulating expression of a THAP responsive gene, said method comprising modulating the interaction of a THAP-family polypeptide or a biologically active fragment thereof with a nucleic acid, thereby enhancing or repressing expression of said THAP responsive gene.
2. The method of Claim 1, wherein said THAP-family polypeptide is THAP1.
3. The method of Claim 1, wherein said nucleic acid is a THAP responsive promoter.
4. The method of Claim 3, wherein said THAP responsive promoter comprises a THAP responsive element.
5. The method of Claim 4, wherein said THAP responsive element is a DR-5 element.
6. The method of Claim 4, wherein said THAP responsive element is an ER-11 element.
7. The method of Claim 4, wherein said THAP responsive element is THRE.
8. The method of Claim 3, wherein said THAP responsive promoter does not comprise a THAP responsive element.
9. The method of Claim 8, wherein said THAP responsive promoter is modulated by a product of a gene that is under the control of a promoter which comprises a THAP responsive element.
10. The method of Claim 1, wherein said THAP responsive gene is selected from the group consisting of Survivin, PTTG1/Securin, PTTG2/Securin, PTTG3/Securin, CKS1, MAD2L1, USP16/Ubp-M, HMMR/RHAMM, KIAA0008/HURP, CDCA7/JPO1 and THAP1.
11. The method of Claim 1, wherein said THAP responsive gene encodes a polypeptide involved in the G2 or M phase of the cell cycle.
12. The method of Claim 1, wherein said THAP responsive gene encodes a polypeptide involved in the S phase of the cell cycle.
13. The method of Claim 12, wherein said THAP responsive gene encodes a polypeptide involved in DNA replication.
14. The method of Claim 12, wherein said THAP responsive gene encodes a polypeptide involved in DNA repair.
15. The method of Claim 1, wherein said THAP responsive gene encodes a polypeptide involved in RNA splicing.
16. The method of Claim 1, wherein said THAP responsive gene encodes a polypeptide involved in apoptosis.
17. The method of Claim 1, wherein said THAP responsive gene encodes a polypeptide involved in angiogenesis.
18. The method of Claim 1, wherein said THAP responsive gene encodes a polypeptide involved in the proliferation of cancer cells.

19. The method of Claim 1, wherein said THAP responsive gene encodes a polypeptide involved in inflammatory disease.
20. A method of modulating the expression of a gene responsive to a THAP/chemokine complex, said method comprising modulating the interaction of a chemokine with a THAP-family polypeptide or a biologically active fragment thereof, thereby enhancing or repressing expression of said gene.
21. The method of Claim 20, wherein said THAP-family polypeptide is THAP1.
22. The method of Claim 20, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.
23. The method of Claim 20, wherein said chemokine is SLC.
24. The method of Claim 20, wherein said chemokine is CXCL9.
25. The method of Claim 20, wherein the interaction between said chemokine and said THAP-family polypeptide is modulated by providing a THAP-type chemokine-binding agent.
26. The method of Claim 25, wherein said THAP-type chemokine-binding agent comprises a polypeptide selected from the group consisting of a THAP1 polypeptide, a chemokine-binding domain of a THAP1 polypeptide, a THAP1 polypeptide oligomer, an oligomer comprising a THAP1 chemokine-binding domain, a THAP1 polypeptide-immunoglobulin fusion, a THAP1 chemokine-binding domain-immunoglobulin fusion and polypeptide homologs of any one of the aforementioned polypeptides.
27. The method of Claim 26, wherein said chemokine-binding domain is an SLC-binding domain.
28. The method of Claim 26, wherein said chemokine-binding domain is a CXCL9-binding domain.
29. The method of Claim 20, wherein said gene encodes a polypeptide involved in the G2 or M phase of the cell cycle.
30. The method of Claim 20, wherein said gene encodes a polypeptide involved in the S phase of the cell cycle.
31. The method of Claim 30, wherein said gene encodes a polypeptide involved in DNA replication.
32. The method of Claim 30, wherein said gene encodes a polypeptide involved in DNA repair.
33. The method of Claim 20, wherein said gene encodes a polypeptide involved in RNA splicing.
34. The method of Claim 20, wherein said gene encodes a polypeptide involved in apoptosis.
35. The method of Claim 20, wherein said gene encodes a polypeptide involved in angiogenesis.

36. The method of Claim 20, wherein said gene encodes a polypeptide involved in the proliferation of cancer cells.

37. The method of Claim 20, wherein said gene encodes a polypeptide involved in inflammatory disease.

5 38. A method of modulating the expression of a gene responsive to a THAP/chemokine complex, said method comprising modulating the interaction of a THAP/chemokine complex with a nucleic acid, thereby enhancing or repressing expression of said gene.

39. The method of Claim 38, wherein said THAP-family polypeptide is THAP1.

40. The method of Claim 38, wherein said chemokine is selected from the group  
10 consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

41. The method of Claim 38, wherein said chemokine is SLC.

42. The method of Claim 38, wherein said chemokine is CXCL9.

43. The method of Claim 38, wherein said gene encodes a polypeptide involved in the G2 or M phase of the cell cycle.

15 44. The method of Claim 38, wherein said gene encodes a polypeptide involved in the S phase of the cell cycle.

45. The method of Claim 44, wherein said gene encodes a polypeptide involved in DNA replication.

46. The method of Claim 44, wherein said gene encodes a polypeptide involved in  
20 DNA repair.

47. The method of Claim 38, wherein said gene encodes a polypeptide involved in RNA splicing.

48. The method of Claim 38, wherein said gene encodes a polypeptide involved in apoptosis.

25 49. The method of Claim 38, wherein said gene encodes a polypeptide involved in angiogenesis.

50. The method of Claim 38, wherein said gene encodes a polypeptide involved in the proliferation of cancer cells.

30 51. The method of Claim 38, wherein said gene encodes a polypeptide involved in inflammatory disease.

52. The method of Claim 38, wherein said nucleic acid is a THAP responsive promoter.

53. The method of Claim 52, wherein said THAP responsive promoter comprises a THAP responsive element.

35 54. The method of Claim 53, wherein said THAP responsive element is a DR-5 element.

55. The method of Claim 53, wherein said THAP responsive element is an ER-11 element.
56. The method of Claim 53, wherein said THAP responsive element is THRE.
57. The method of Claim 52, wherein said THAP responsive promoter does not  
5 comprise a THAP responsive element.
58. The method of Claim 57, wherein said THAP responsive promoter is modulated by a product of a gene that is under the control of a promoter which comprises a THAP responsive element.
59. A pharmaceutical composition comprising a THAP responsive element in a  
10 pharmaceutically acceptable carrier.
60. The pharmaceutical composition of Claim 59, wherein said THAP responsive element is a DR-5 element.
61. The pharmaceutical composition of Claim 59, wherein said THAP responsive element is an ER-11 element.
- 15 62. The pharmaceutical composition of Claim 59, wherein said THAP responsive element is an THRE.
63. A transcription factor decoy consisting essentially of a THAP responsive element.
64. The transcription factor decoy of Claim 63, wherein said THAP responsive element is a DR-5 element.
- 20 65. The transcription factor decoy of Claim 63, wherein said THAP responsive element is a ER-11 element.
66. The transcription factor decoy of Claim 63, wherein said THAP responsive element is a THRE element.
67. A cell comprising a transcription factor decoy of claim 63.
- 25 68. A method of modulating the interaction between a nucleic acid and a THAP-family polypeptide or a biologically active fragment thereof, said method comprising providing a transcription factor decoy which comprises a THAP responsive element, thereby modulating the interaction between said nucleic acid and said THAP-family polypeptide or a biologically active fragment thereof.
- 30 69. The method of Claim 68, wherein said THAP-family polypeptide is THAP1.
70. The method of Claim 68, wherein said THAP responsive element is a DR-5 element.
71. The method of Claim 68, wherein said THAP responsive element is an ER-11 element.
- 35 72. The method of Claim 68, wherein said THAP responsive element is THRE.
73. A method of modulating the interaction between a nucleic acid and a THAP/chemokine complex, said method comprising providing a transcription factor decoy which



comprises a THAP responsive element, thereby modulating the interaction between said nucleic acid and said THAP/chemokine complex.

74. The method of Claim 73, wherein said THAP-family polypeptide is THAP1.
75. The method of Claim 73, wherein said chemokine is selected from the group  
5 consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.
76. The method of Claim 73, wherein said chemokine is SLC.
77. The method of Claim 73, wherein said chemokine is CXCL9.
78. The method of Claim 73, wherein said THAP responsive element is a DR-5  
element.
- 10 79. The method of Claim 73, wherein said THAP responsive element is an ER-11  
element.
80. The method of Claim 73, wherein said THAP responsive element is THRE.
81. A vector packaging cell line comprising a cell comprising a viral vector which  
comprises a promoter operably linked to a nucleic acid encoding a THAP-family polypeptide or a  
15 biologically active fragment thereof.
82. The cell line of Claim 81, wherein said cell further comprises an introduced nucleic  
acid construct comprising a nucleic acid encoding a chemokine operably linked to a promoter.
83. The cell line of Claim 82, wherein said chemokine-encoding construct is included  
on the same vector as said nucleic acid encoding said THAP-family polypeptide or biologically  
20 active fragment thereof.
84. The cell line of Claim 82, wherein said nucleic acid encoding said chemokine  
encodes a chemokine selected from the group consisting of SLC, CCL19, CCL5, CXCL11,  
CXCL10 and CXCL9.
85. The cell line of Claim 82, wherein said nucleic acid encoding said chemokine  
25 encodes SLC.
86. The cell line of Claim 82, wherein said nucleic acid encoding said chemokine  
encodes CXCL9.
87. The cell line of Claim 81, wherein said THAP-family polypeptide is THAP1.
88. The cell line of Claim 81, wherein said cell is a mammalian cell.
- 30 89. The cell line of Claim 88, wherein said cell is a human cell.
90. The cell line of Claim 81, wherein said viral vector is an adenoviral vector.
91. The cell line of Claim 81, wherein said viral vector is a retroviral vector.
92. A cell which is genetically engineered to express a THAP-family polypeptide or a  
biologically active fragment thereof.
- 35 93. The cell line of Claim 92, wherein said THAP-family polypeptide is THAP1.
94. The cell line of Claim 92, wherein said cell is a mammalian cell.
95. The cell line of Claim 92, wherein said cell is a human cell.

96. The cell line of Claim 92, wherein said THAP family polypeptide is encoded by a gene that is introduced into the cell on an adenoviral vector.

97. The cell line of Claim 92, wherein said THAP family polypeptide is encoded by a gene that is introduced into the cell on a retroviral vector.

5 98. A method of constructing a cell which expresses a recombinant THAP-family polypeptide, said method comprising introducing into a cell a vector comprising a nucleic acid encoding a THAP-family polypeptide or a biologically active fragment thereof operably linked to a promoter.

99. The method of Claim 98, further comprising introducing into a cell a nucleic acid  
10 construct comprising a nucleic acid encoding a chemokine operably linked to a promoter.

100. The method of Claim 99, wherein said chemokine-encoding construct is included on the same vector as said nucleic acid encoding said THAP-family polypeptide or biologically active fragment thereof.

101. The method of Claim 99, wherein said nucleic acid encoding said chemokine  
15 encodes a chemokine selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

102. The method of Claim 99, wherein said nucleic acid encoding said chemokine encodes SLC.

103. The method of Claim 99, wherein said nucleic acid encoding said chemokine  
20 encodes CXCL9.

104. The method of Claim 98, wherein said THAP-family polypeptide is THAP1.

105. The method of Claim 98, wherein said cell is a mammalian cell.

106. The method of Claim 105, wherein said cell is a human cell.

107. The method of Claim 98, wherein said vector is a viral vector.

25 108. The method of Claim 107, wherein said vector is an adenoviral vector.

109. The method of Claim 107, wherein said vector is a retroviral vector.

110. The method of Claim 98, wherein said vector is introduced into said cell by transfection.

111. A method of ameliorating symptoms associated with a condition mediated by a  
30 THAP/chemokine complex, said method comprising:

introducing into a cell a nucleic acid construct comprising a nucleic acid encoding a chemokine operably linked to a promoter and a nucleic acid construct comprising a nucleic acid encoding a THAP-family polypeptide or a biologically active fragment thereof operably linked to a promoter; and

35 expressing said nucleic acid encoding said chemokine and said nucleic acid encoding said THAP-family polypeptide or biologically active fragment thereof.

112. The method of Claim 111, wherein said nucleic acid constructs are present on a single vector.

113. The method of Claim 111, wherein said nucleic acid constructs are present on different vectors.

5 114. The method of Claim 111, wherein said cell is a mammalian cell.

115. The method of Claim 114, wherein said cell is a human cell.

116. The method of Claim 111, wherein said nucleic acid encoding said chemokine encodes a chemokine selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

10 117. The method of Claim 111, wherein said nucleic acid encoding said chemokine encodes SLC.

118. The method of Claim 111, wherein said nucleic acid encoding said chemokine encodes CXCL9.

119. The method of Claim 111, wherein said THAP-family polypeptide is THAP1.

15 120. A method of identifying a test compound that modulates transcription at a THAP responsive element, said method comprising:

comparing the level of transcription from a THAP responsive promoter in the presence and absence of a test compound wherein a determination that the level of transcription is increased or decreased in the presence of said test compound relative to the level of transcription in the absence of said test compound indicates that said test compound is a candidate modulator of transcription.

20 121. The method of Claim 120, wherein the level of transcription from said THAP responsive promoter in the presence and absence of the test compound is determined by performing an in vitro transcription reaction using a construct comprising said THAP responsive promoter and a THAP-family polypeptide or a biologically active fragment thereof, wherein said THAP-family polypeptide comprises an amino acid sequence having at least 30% amino acid identity to an amino acid sequence of SEQ ID NO: 1.

25 122. The method of Claim 120, wherein the level of transcription from said THAP responsive promoter in the presence and the absence of the test compound is determined by measuring the level of transcription from a THAP responsive promoter in a cell expressing a THAP-family polypeptide or a biologically active fragment thereof, wherein said THAP-family polypeptide comprises an amino acid sequence having at least 30% amino acid identity to an amino acid sequence of SEQ ID NO: 1.

30 123. The method of Claim 120, wherein said THAP-family polypeptide or biologically active fragment thereof is selected from the group consisting of SEQ ID NOs: 1-114 and biologically active fragments thereof.

124. The method of Claim 120, wherein said THAP responsive promoter comprises a THAP responsive element having a nucleotide sequence selected from the group consisting of SEQ ID NOs: 140-159, SEQ ID NO: 306, and homologs thereof having at least 60% nucleotide identity.

5 125. The method of Claim 121 or Claim 122, wherein the level of transcription in the presence or absence of said test compound is measured in the presence of a chemokine.

126. The method of Claim 125, wherein said chemokine is selected from the group consisting of CCL family chemokines and CXCL family chemokines.

127. The method of Claim 126, wherein said CCL family chemokine is selected from the group consisting of SLC, CCL19 and CCL5.

10 128. The method of Claim 126, wherein said CXCL family chemokine is selected from the group consisting of CXCL11, CXCL10 and CXCL9.

129. The method of Claim 125, wherein the level of transcription in the presence or absence of said test compound is measured in a cell which expresses a receptor for said chemokine.

15 130. The method of Claim 129, wherein said chemokine receptor is selected from the group consisting of CCR1, CCR3, CCR5, CCR7, CCR11 and CXCR3.

131. The method of Claim 130, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

132. The method of Claim 129, wherein said THAP-family polypeptide comprises THAP1 or a biologically active fragment thereof and said cell expresses the CCR7 receptor.

20 133. The method of Claim 132, wherein said chemokine is SLC.

134. The method of Claim 129, wherein said THAP-family polypeptide comprises THAP1 or a biologically active fragment thereof and said cell expresses the CXCR3 receptor.

135. The method of Claim 134, wherein said chemokine is CXCL9.

25 136. The method of Claim 122, wherein said THAP responsive promoter is in a gene endogenous to said cell.

137. The method of Claim 122, wherein said THAP responsive promoter has been introduced into said cell.

138. The method of Claim 122, wherein said THAP responsive promoter does not comprise a THAP responsive element.

30 139. The method of Claim 138, wherein said THAP responsive promoter is modulated by a product of a gene that is under the control of a promoter which comprises a THAP responsive element.

35 140. A method for reducing the symptoms associated with a condition selected from the group consisting of excessive or insufficient angiogenesis, inflammation, metastasis of a cancerous tissue, excessive or insufficient apoptosis, cardiovascular disease and neurodegenerative diseases comprising modulating the interaction between a THAP-family polypeptide and a chemokine in an individual suffering from said condition.

141. The method of Claim 140, wherein said THAP-family polypeptide is selected from a group consisting of polypeptides having an amino acid sequence of SEQ ID NOs: 1-114.

142. The method of Claim 140, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

5 143. The method of Claim 140, wherein said chemokine is SLC and the condition is inflammation.

144. The method of Claim 140, wherein said chemokine is SLC and the condition is excessive or insufficient angiogenesis.

10 145. The method of Claim 140, wherein said chemokine is CXCL9 and the condition is inflammation.

146. The method of Claim 140, wherein said chemokine is CXCL9 and the condition is excessive or insufficient angiogenesis.

147. A method for reducing the symptoms associated with a condition resulting from the activity of a chemokine in an individual comprising modulating the interaction between said  
15 chemokine and a THAP-family polypeptide in said individual.

148. The method of Claim 147, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

149. The method of Claim 147, wherein said chemokine is SLC.

150. The method of Claim 147, wherein said chemokine is CXCL9.

20 151. The method of Claim 147, wherein said THAP-family polypeptide is THAP-1.

152. The method of Claim 147, wherein the condition is inflammation.

153. The method of Claim 147, wherein the condition is excessive or insufficient angiogenesis.

154. The method of Claim 147, wherein the interaction between said chemokine and  
25 said THAP-family polypeptide is modulated by administering to an individual, a therapeutically effective amount of a THAP-type chemokine-binding agent.

155. The method of Claim 154, wherein said THAP-type chemokine-binding agent comprises a therapeutically effective amount of a polypeptide selected from the group consisting of a THAP1 polypeptide, an chemokine-binding domain of a THAP1 polypeptide, a THAP1  
30 polypeptide oligomer, an oligomer comprising a THAP1 chemokine-binding domain, a THAP1 polypeptide-immunoglobulin fusion, a THAP1 chemokine-binding domain-immunoglobulin fusion and polypeptide homologs having at least 30% amino acid identity to any one of the aforementioned polypeptides.

35 156. The method of Claim 155, wherein said chemokine-binding domain is an SLC-binding domain.

157. The method of Claim 155, wherein said chemokine-binding domain is a CXCL9-binding domain.

158. A method of reducing the symptoms associated with a condition resulting from the activity of a THAP-family polypeptide in an individual comprising modulating the extent of transcriptional repression or activation of at least one THAP-family responsive promoter in said individual.

5 159. The method of Claim 158, wherein said THAP-family polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114.

160. The method of Claim 158, wherein said THAP-family polypeptide comprises an amino acid sequence of SEQ ID NO: 3.

10 161. The method of Claim 158, wherein said THAP responsive promoter comprises a THAP responsive element.

162. The method of Claim 158, wherein said THAP responsive promoter does not comprise a THAP responsive element.

163. A method of reducing the symptoms associated with a condition resulting from the activity of a THAP-family polypeptide in an individual, said method comprising:  
15 diagnosing said individual with a condition resulting from the activity of a THAP-family polypeptide; and  
administering a compound which modulates the interaction between said THAP-family polypeptide and a chemokine to said individual.

20 164. The method of Claim 163, wherein said THAP-family polypeptide is selected from a group consisting of polypeptides having an amino acid sequence of SEQ ID NOs: 1-114.

165. The method of Claim 163, wherein said THAP-family polypeptide is THAP1.

166. The method of Claim 163, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

167. The method of Claim 163, wherein said chemokine is SLC.

25 168. The method of Claim 163, wherein said chemokine is CXCL9.

169. A method of reducing the symptoms associated with a condition resulting from the activity of a THAP-family polypeptide in an individual comprising:

30 diagnosing said individual with a condition resulting from the activity of THAP-family polypeptide; and  
administering a chemokine or an analog thereof to said individual.

170. The method of Claim 169, wherein said THAP-family polypeptide is selected from a group consisting of polypeptides having an amino acid sequence of SEQ ID NOs: 1-114.

171. The method of Claim 169, wherein said THAP-family polypeptide is THAP1.

35 172. The method of Claim 169, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

173. The method of Claim 169, wherein said chemokine is SLC.

174. The method of Claim 169, wherein said chemokine is CXCL9.

175. A method of reducing the symptoms associated with transcriptional repression or activation mediated by a THAP-family polypeptide in an individual comprising administering a chemokine or an analog thereof to said individual.

176. The method of Claim 175, wherein said THAP-family polypeptide is selected from  
5 a group consisting of polypeptides having an amino acid sequence of SEQ ID NOs: 1-114.

177. The method of Claim 175, wherein said THAP-family polypeptide is THAP1.

178. The method of Claim 175, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

179. The method of Claim 175, wherein said chemokine is SLC.

10 180. The method of Claim 175, wherein said chemokine is CXCL9.

181. A method of reducing the symptoms associated with the activity of a chemokine in an individual comprising modulating the extent to which said chemokine is transported to the nucleus of a cell in said individual.

182. The method of Claim 181, wherein said chemokine is selected from the group  
15 consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

183. The method of Claim 181, wherein said cell expresses a chemokine receptor selected from the group consisting of CCR1, CCR3, CCR5, CCR7, CCR11 and CXCR3.

184. The method of Claim 183, wherein said chemokine is SLC and said chemokine receptor is CCR7.

20 185. The method of Claim 183, wherein said chemokine is CXCL9 and said chemokine receptor is CXCR3.

186. The method of Claim 181, wherein the extent of transport of said chemokine into a nucleus of a cell is modulated by contacting said chemokine with a THAP-type chemokine-binding agent.

25 187. The method of Claim 186, wherein said THAP-type chemokine-binding agent selected from the group consisting of a THAP1 polypeptide, a chemokine-binding domain of a THAP1 polypeptide, a THAP1 polypeptide oligomer, an oligomer comprising a THAP1 chemokine-binding domain, a THAP1 polypeptide-immunoglobulin fusion, a THAP1 chemokine-binding domain-immunoglobulin fusion and polypeptide homologs having at least 30% amino acid  
30 identity to any one of the aforementioned polypeptides.

188. The method of Claim 187, wherein said chemokine-binding domain is an SLC-binding domain.

189. The method of Claim 187, wherein said chemokine-binding domain is a CXCL9-binding domain.

35 190. A method for identifying a compound which modulates the transport of a chemokine into the nucleus comprising comparing the extent of said chemokine transport into the nucleus of cells in the presence and absence of a test compound.

191. The method of Claim 190, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

192. The method of Claim 190, wherein said cell expresses a chemokine receptor selected from the group consisting of CCR1, CCR3, CCR5, CCR7, CCR11 and CXCR3.

5 193. The method of Claim 192, wherein said chemokine is SLC and said chemokine receptor is CCR7.

194. The method of Claim 192, wherein said chemokine is CXCL9 and said chemokine receptor is CXCR3.

10 195. The method of Claim 190, wherein the extent of transport of said chemokine into a nucleus of a cell is modulated by contacting said chemokine with a THAP-type chemokine-binding agent.

196. The method of Claim 195, wherein said THAP-type chemokine-binding agent is selected from the group consisting of a THAP1 polypeptide, a chemokine-binding domain of a THAP1 polypeptide, a THAP1 polypeptide oligomer, an oligomer comprising a THAP1 chemokine-binding domain, a THAP1 polypeptide-immunoglobulin fusion, a THAP1 chemokine-binding domain-immunoglobulin fusion and polypeptide homologs having at least 30% amino acid identity to any one of the aforementioned polypeptides.

197. The method of Claim 196, wherein said chemokine-binding domain is an SLC-binding domain.

20 198. The method of Claim 196, wherein said chemokine-binding domain is a CXCL9-binding domain.

199. The method of Claim 190, wherein transport of SLC into the nucleus is measured by immunostaining.

25 200. A vector comprising a THAP responsive promoter operably linked to a nucleic acid encoding a detectable product.

201. The vector of Claim 200, wherein said THAP responsive promoter comprises a THAP responsive element.

202. The vector of Claim 200, wherein said THAP responsive promoter does not comprise a THAP responsive element.

30 203. A genetically engineered cell comprising the vector of any one of Claims 200-202.

204. An *in vitro* transcription reaction comprising a nucleic acid comprising a THAP responsive promoter, ribonucleotides and an RNA polymerase.

205. The *in vitro* transcription reaction of Claim 204, wherein said THAP responsive promoter comprises a THAP responsive element.

35 206. An isolated mutant THAP-family polypeptide that does not bind to a chemokine.



207. The isolated mutant THAP-family polypeptide of Claim 206, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL11, CXCL10 and CXCL9.

5 208. The isolated mutant THAP-family polypeptide of Claim 206, wherein said chemokine is SLC.

209. The isolated mutant THAP-family polypeptide of Claim 206, wherein said chemokine is CXCL9.

210. The isolated mutant THAP-family polypeptide of Claim 206, wherein said THAP-family polypeptide is THAP1.

10 211. The isolated mutant THAP-family polypeptide of Claim 210, wherein said polypeptide comprises an amino acid sequence of SEQ ID NO: 3.

212. The isolated mutant THAP-family polypeptide of Claim 211, wherein said amino acid sequence comprises at least one point mutation.

FIGURE 1

**A**

hTHAP1	1	MVQSCSAYGCKNRYDKDKPVSFHKFPLTRPSLCKEWEAAVRRKNFKPTKYSSICSEHFTP
mTHAP1	1	MVQSCSAYGCKNRYDKDKPVSFHKFPLTRPSLCKEWEAAVRRKNFKPTKYSSICSEHFTP
hTHAP1	61	DCFKREGCNKLLKENAVPTIFLCTEPHDKKEDLLEPQQLPPPLPPVPSQVDAAGLLM
mTHAP1	61	DCFKREGCNKLLKENAVPTIFLYIEPHEKKEDL-ESQEQLPSPS--PPASQVDAAGLLM
hTHAP1	121	PPLQTPVNLVSFCDHNYTVEDTMHQKRRIHOLEQQVEKLRKRLKTAQQRCRRQERQLEKL
mTHAP1	118	PPLQTPDNLSVFCDHNYTVEDTMHQKRRIHOLEQQVEKLRKRLKTAQQRCRRQERQLEKL
hTHAP1	181	KEVVFQKEKDDVSERGVILENDYFEIVEVPA
mTHAP1	178	KEVVFQREKDDASERGVILENDYFEIVEVPA

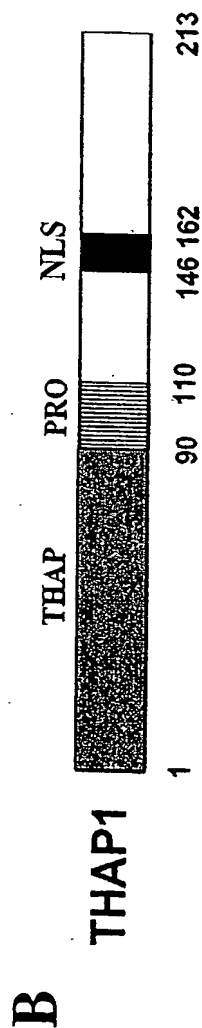


FIGURE 2

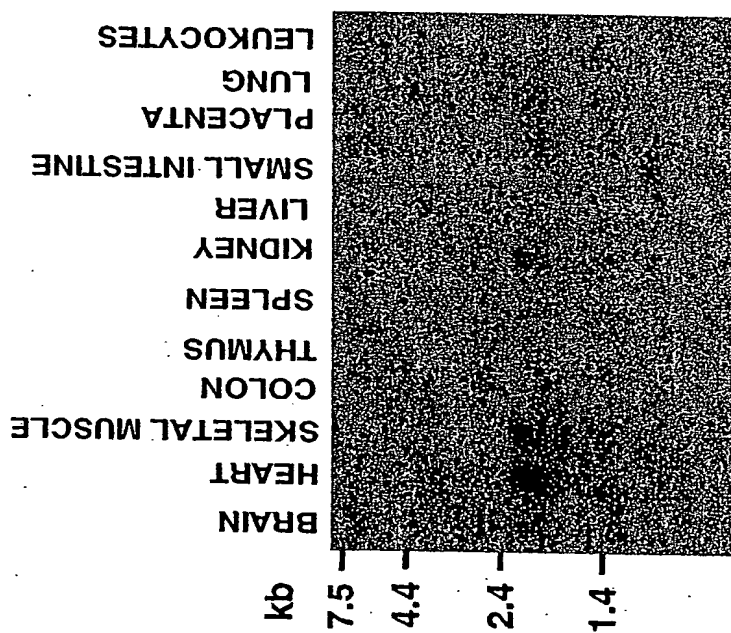


Figure 3

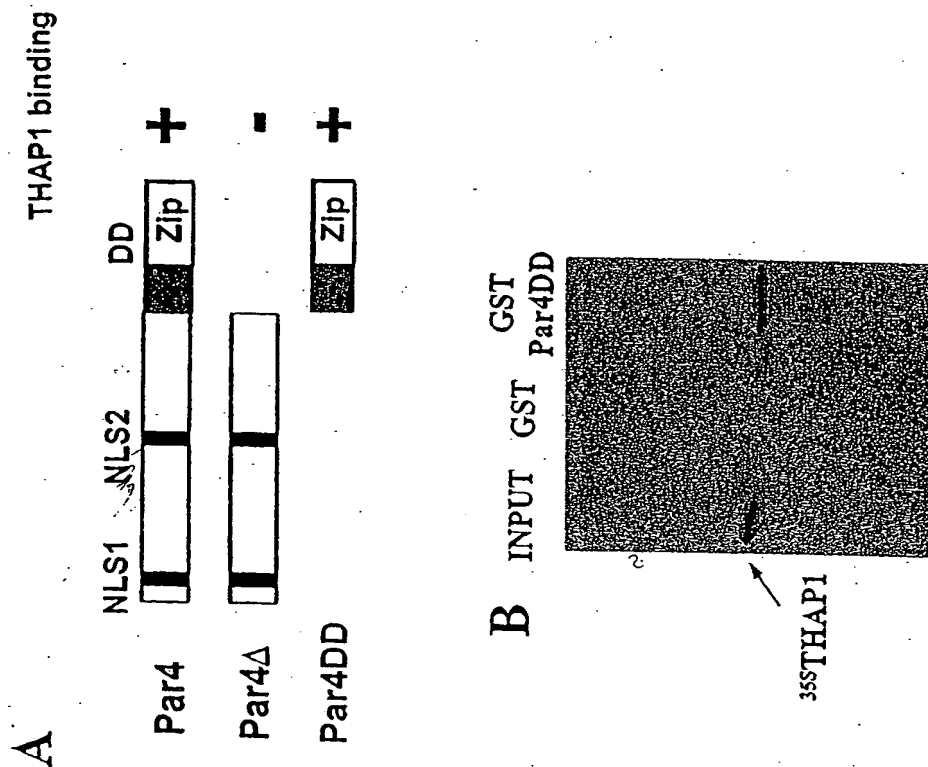


FIGURE 4A

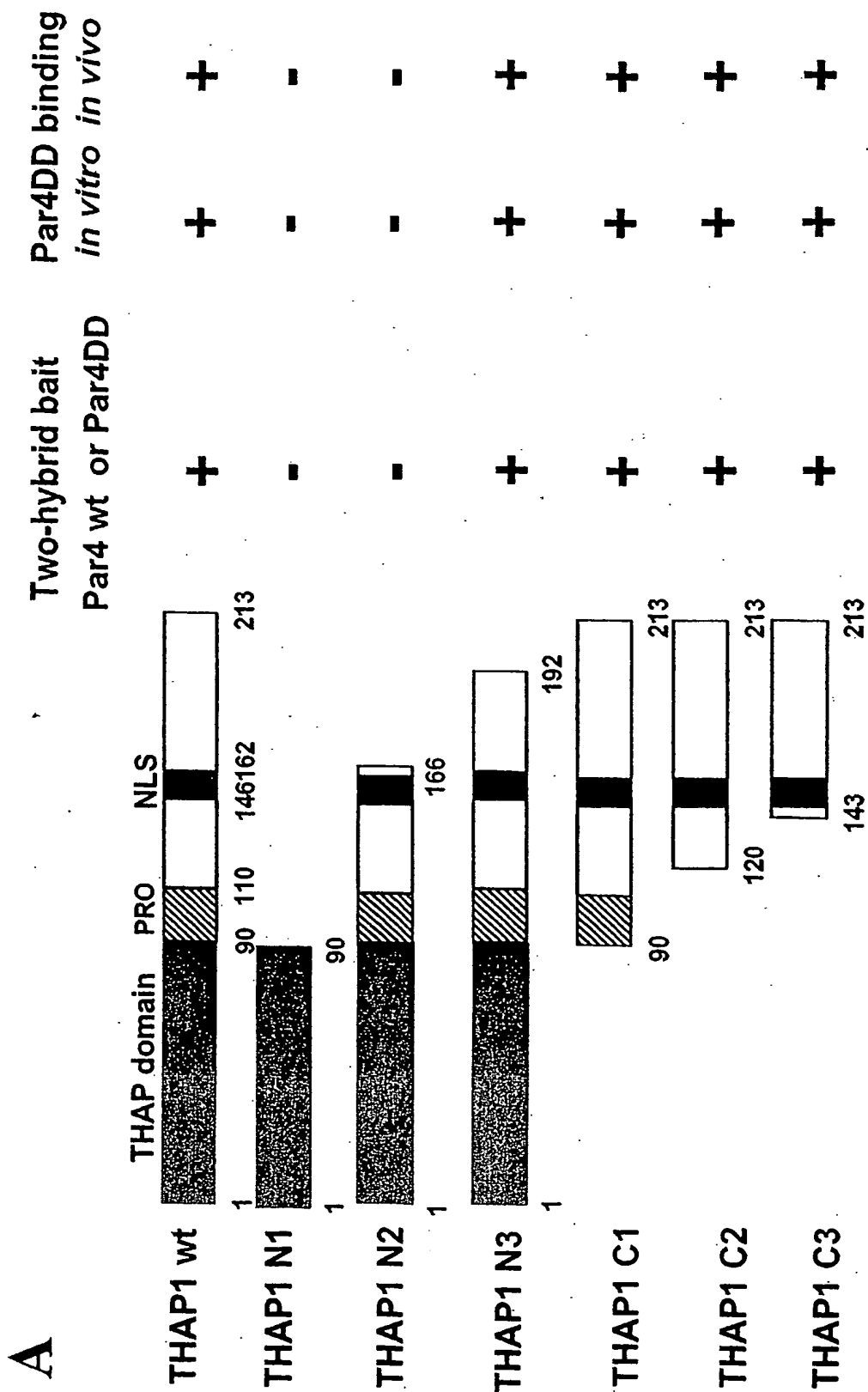


Figure 4b

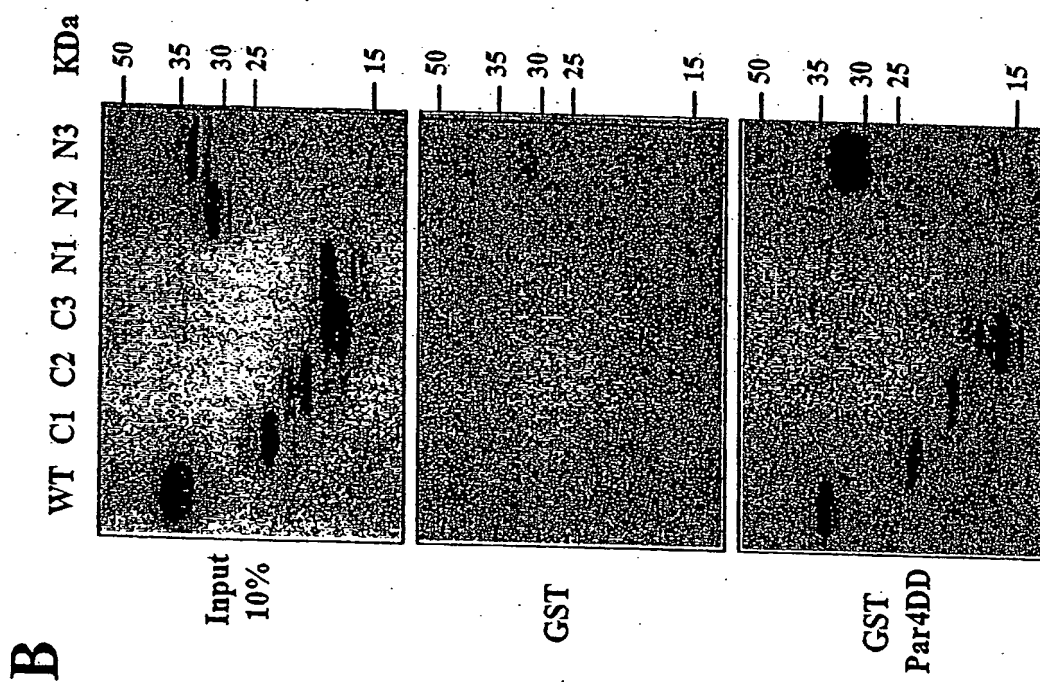


Figure 5

A

mZIP -VLEDVAAAEQGLREL--QRGRQCRERVCAALRAAEQREARCRDG  
 mTHAP-1 -QLEQQVEKLKKLKTAAQQRRCRQERQLEKLKEVVHFQREKDDASE  
 hTHAP-1 -QLEQQVEKLKKLKTAAQQRRCRQERQLEKLKEVVHFQREKDDVSE

Consensus Par4 binding site: LE (X<sub>12-14</sub>) QRXRRQXR (X<sub>1</sub>) QXE

B

Two-hybrid bait  
 Par4 wt or Par4DD *in vitro* *in vivo*

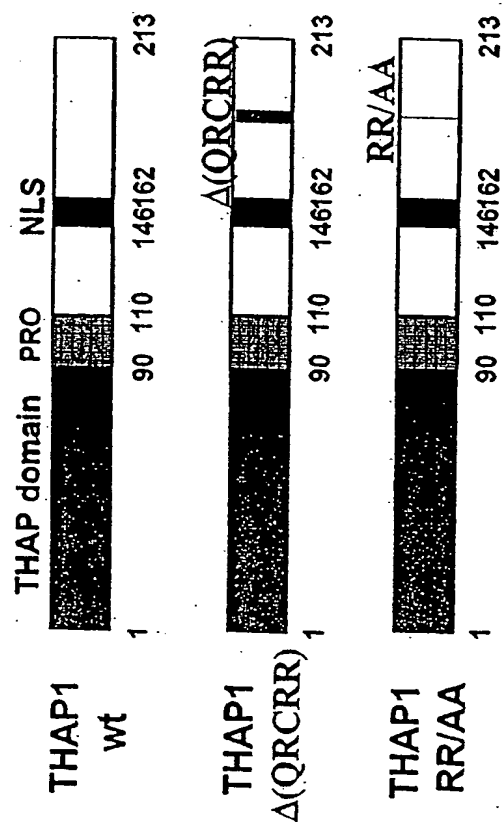


Figure 6

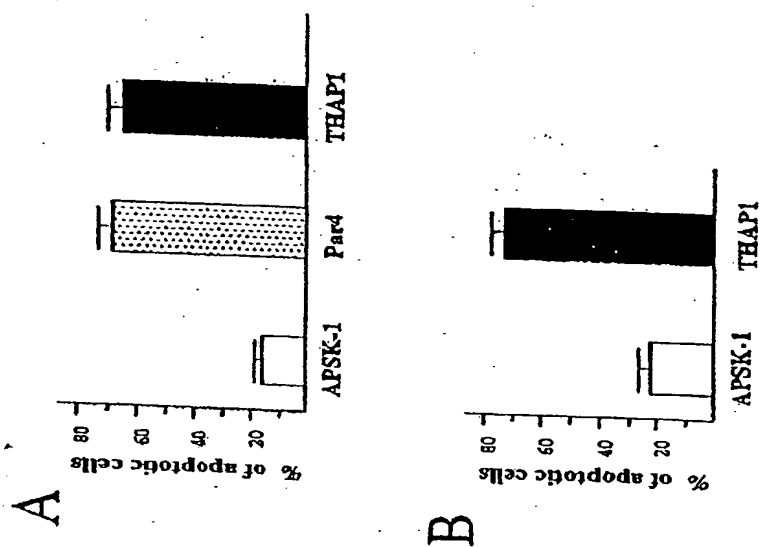
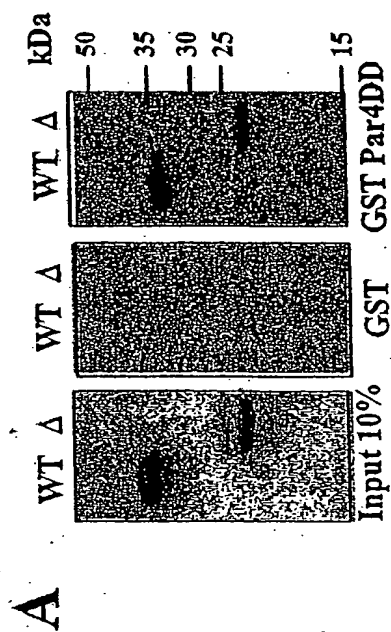




Figure 7



**B**

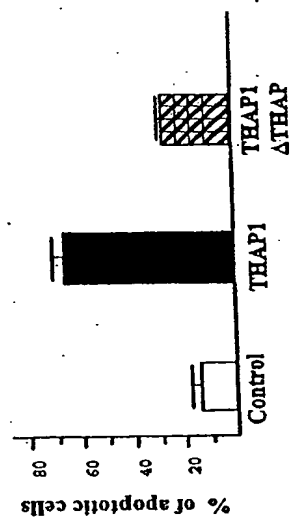


Figure 8

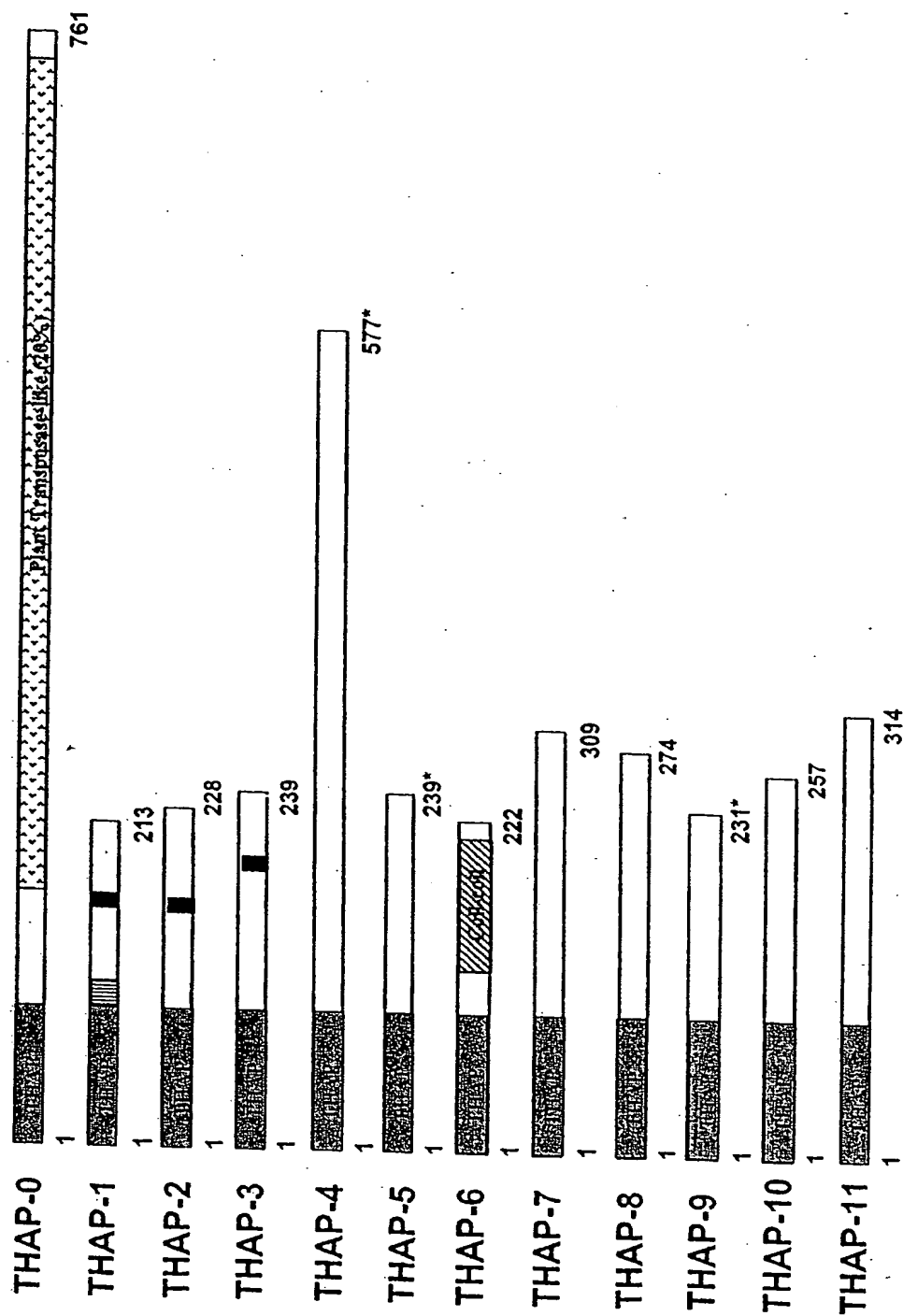


Figure 9

A

hTHAP1 1 MVQCSAAGCKNRDYDKDPVSPHKKFPTTRPSLCKEEMEAARRKNFKPTKYSSICSEHFTP  
dmTransposase 1 -MKYIC-KECCN-AVIGVALLHVFKCARR-----KLMBOSIG---CSLGENSQICDTHEND  
consensus 1 mv Cs y Ckn K v K 1 Rpslck WE v rkn S IC HF

hTHAP1 61 DCEK-----RECNNLAKENAVPTIFCTEHHDK  
dmTransposase 51 SQKKAAPAKGOTFKRRRLNADAVPKVIEPEDEKI  
consensus 61 fkaapag k L AVPT 1 EP

B

hTHAP1 1 MVQS-SNYCKKMYD---DCEVSRNKEALTRPSLCKEEMEAR--RKN-----RK  
hTHAP3 1 MPIS-SANRCCKMYSS---DQCEFHRRPSPRELCKEALNIG--RGN-----RK  
hTHAP5 1 MPRK-SANICRRGRNKD-RKSHYFETLHDEKLEKRNK--RDS-----RV  
hTHAP8 1 MPRK-SANICRRGRNKD-RKSHYFETLHDEKLEKRNK--RDS-----RV  
hTHAP8 1 MPRK-SANICRRGRNKD-RKSHYFETLHDEKLEKRNK--RDS-----RV  
hTHAP4 1 MPRK-SANICRRGRNKD-RKSHYFETLHDEKLEKRNK--RDS-----RV  
hTHAP2 1 MPRK-SANICRRGRNKD-RKSHYFETLHDEKLEKRNK--RDS-----RV  
hTHAP0 1 MPRK-SANICRRGRNKD-RKSHYFETLHDEKLEKRNK--RDS-----RV  
hTHAP7 1 MPRK-SANICRRGRNKD-RKSHYFETLHDEKLEKRNK--RDS-----RV  
hTHAP9 1 MPRK-SANICRRGRNKD-RKSHYFETLHDEKLEKRNK--RDS-----RV  
hTHAP6 1 MPRK-SANICRRGRNKD-RKSHYFETLHDEKLEKRNK--RDS-----RV  
hTHAP11 1 MPRK-SANICRRGRNKD-RKSHYFETLHDEKLEKRNK--RDS-----RV  
hTHAP10 1 MPRK-SANICRRGRNKD-RKSHYFETLHDEKLEKRNK--RDS-----RV  
dmTransposase 1 MPRK-SANICRRGRNKD-RKSHYFETLHDEKLEKRNK--RDS-----RV  
consensus 1 mpk C a C nr k k vshxrp hd rr mv v f w

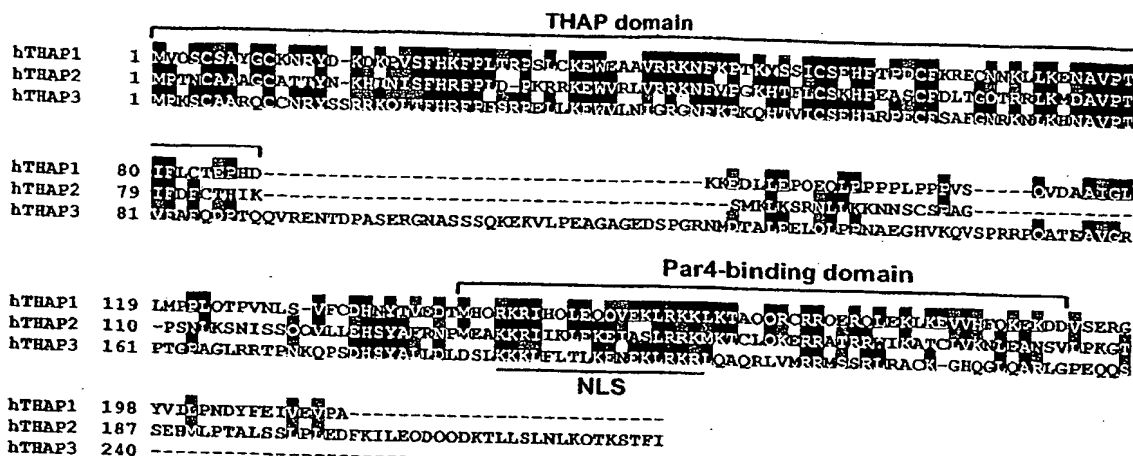
hTHAP1 47 P-----TKYSSICSEHFTDCKRECN-----NELLKENAVPTIFCTEHPDK--  
hTHAP3 48 P-----KQHTVICSSEHFTPECSAFGN-----RMLKHNAVPTVAFQDPTQ--  
hTHAP5 50 P-----SKYQFICSSEHFTDSDIRMG-----RYLQOTAVPTIFS-LPEDN--  
hTHAP8 51 P-----SCHQFICSSEHFTPSGQWRMG-----VRLPDAVPTIISRGPPAK--  
hTHAP4 49 P-----TKYFICSSEHFTKDSKRLD-----OHLNPTAVPTIIS--LTK--  
hTHAP2 46 P-----GKHTFICSSEHFTSCDITQ-----TR-RLMDAVPTIDFCTHIKSM-  
hTHAP0 45 P-----DKTPDLNKHVLCSEHFTSMICTSP-----YRTVMDNPTIIDLTS--  
hTHAP7 55 P-----PGAI-LCSSEHFTSESYGIR-----GRLKRGAVPTISLYK--  
hTHAP9 55 P-----SEYIFCSSEHFTEDCELVGIS-----GRLKRGAVPTISLYK--  
hTHAP6 55 P-----KGV-LCSSEHFTKDEDSAPN-----IKLPGVPTIISIDSYHLO--  
hTHAP11 54 P-----TGHRVLCSEHFTVGG-----GRTYVPTIISIDSYHLO--  
hTHAP10 46 G-----NDRSVICSSEHFTAPACDVSIVI-----QRLFSQRLKAGAVPTI--  
dmTransposase 38 G-----ENSLQDTHENDSQKAAAPAKGOTFKRRRLNADAVPKVIEPEDEKI--  
consensus 61 P fcs hf f k k avpt

BNSDOCID: <WO\_\_\_\_\_2004055050A2 I\_>

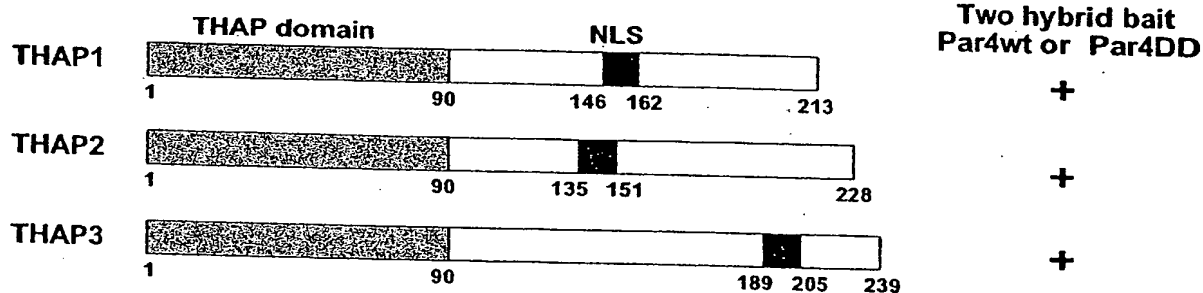
FIG. 9C

2EM101798	1	MDPDAAYGS	NR	TKK	---	LDKGTTHRR	D	VKQ	---	TLAPN	DV	PKPR	---	SUTCS	CH	RPDS	RTGQ	---	TVNR	GV	TEST	NF	SNPLSKLS	---	93
ZBG799610	1	NP	VCS	AYCK	---	KSPSE	YKZA	YKGE	SEHKEP	---	EDGL	YK	---	SVKCS	HE	CKDE	GVGS	---	HGR	ER	SAVPE	TF	FPKLOMKV	---	97
ZAG172827	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	90
ZAI723023	1	MNSIS	KL	LR	CA	SVY	CV	YFCK	---	ISSRN	---	SVSE	SEHKEP	---	D	ATRV	---	SVKCS	HE	CKDE	GVGS	---	FPKLOMKV	---	107
ZAG421661	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	89
ZBG799617	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	105
ZAG422263	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	96
ZEM033392	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	91
ZAG280314	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	84
OTRAP0	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	95
CAV670865	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	90
OBJ003300	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	94
dmCG10431	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	92
dmCG13894	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	90
dmCG6689	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	104
dmLD47616p	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	96
dmD1P2	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	96
dmTRP	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	87
ABM608335	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	92
ABM576425	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	108
ABM583369	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	95
ABM633152	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	100
AAJ281288	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	95
ABM613843	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	107
ABM645725	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	104
ABM584662	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	97
bmAV401481	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	92
bmAV401484	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	96
ceNP_508983	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	89
ceNP_506277	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	100
ceNP_495084	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	86
ceNP_498747.1	1	NP	VCS	AYCK	---	NTLR	---	NKSS	SEHKEP	---	KOPSL	YK	---	SVKCS	HE	CKDE	GVGS	---	KTR	ER	SAVPE	TF	FPKLOMKV	---	97
consensus	1	mp	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	

A



B



C

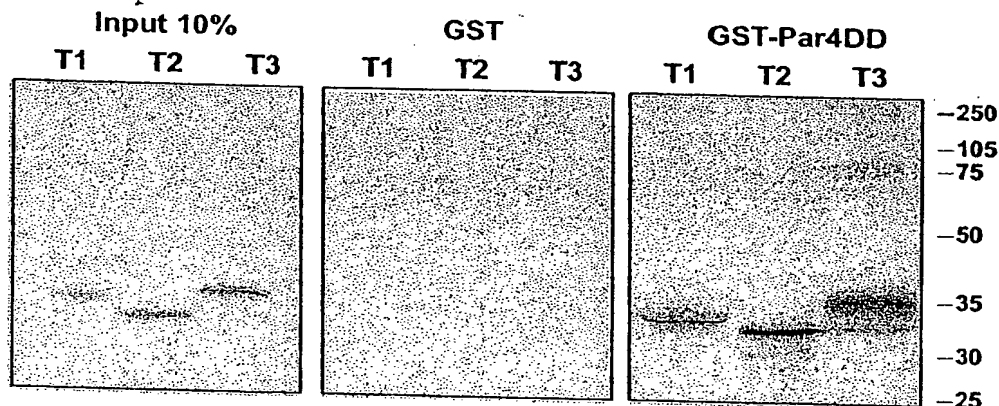


Fig. 14/3611

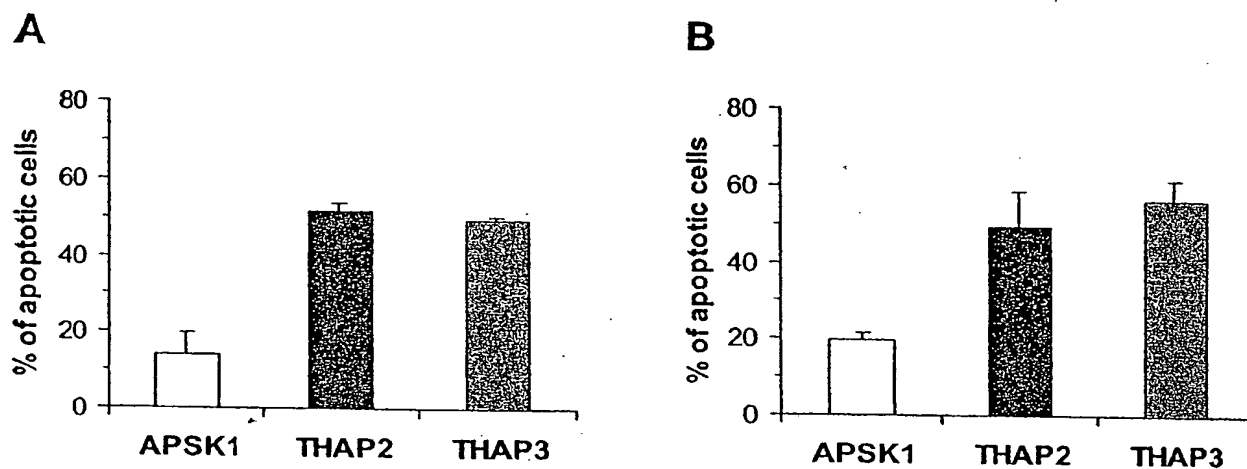


Figure 12

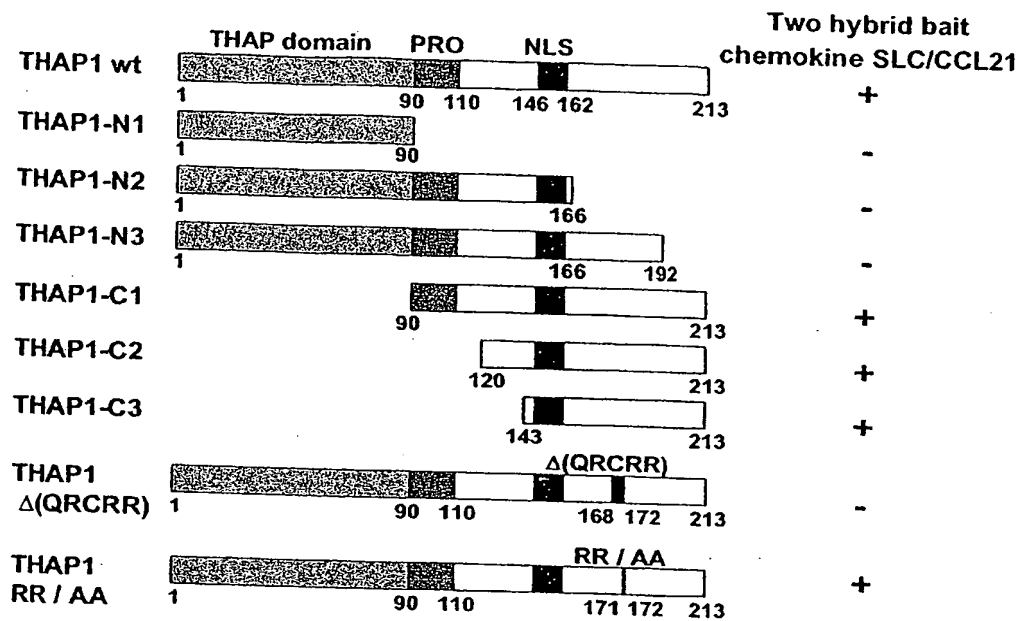




Figure 13

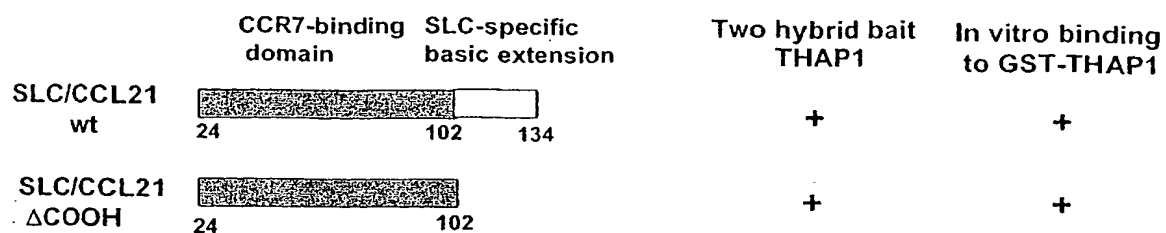


FIGURE 14

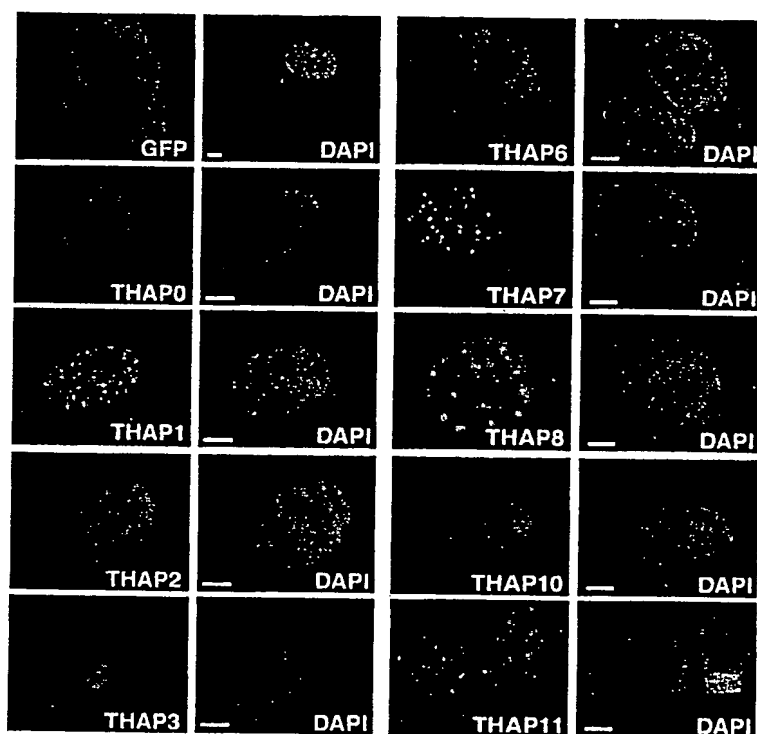


FIGURE 15

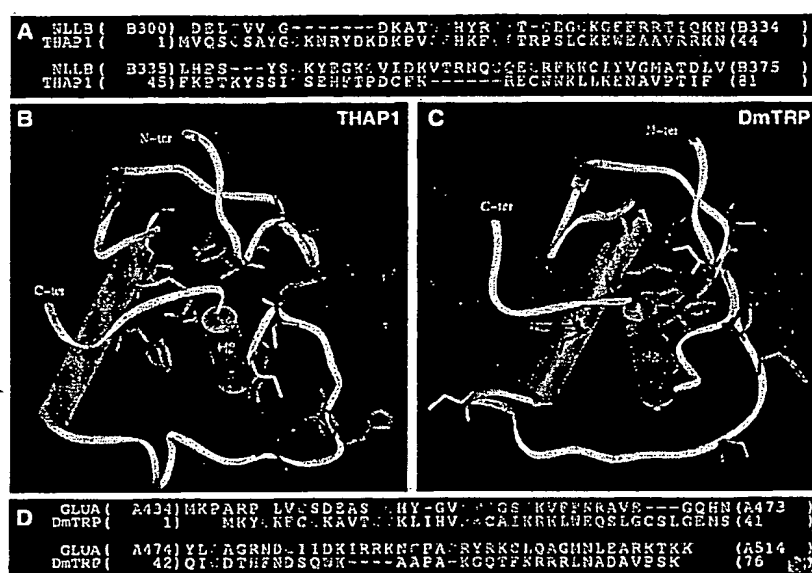
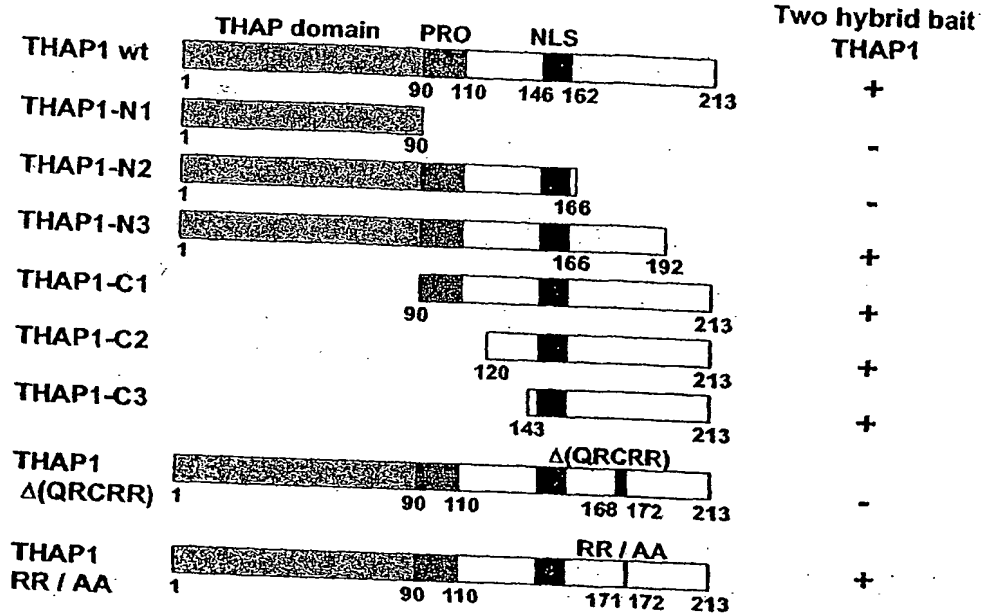
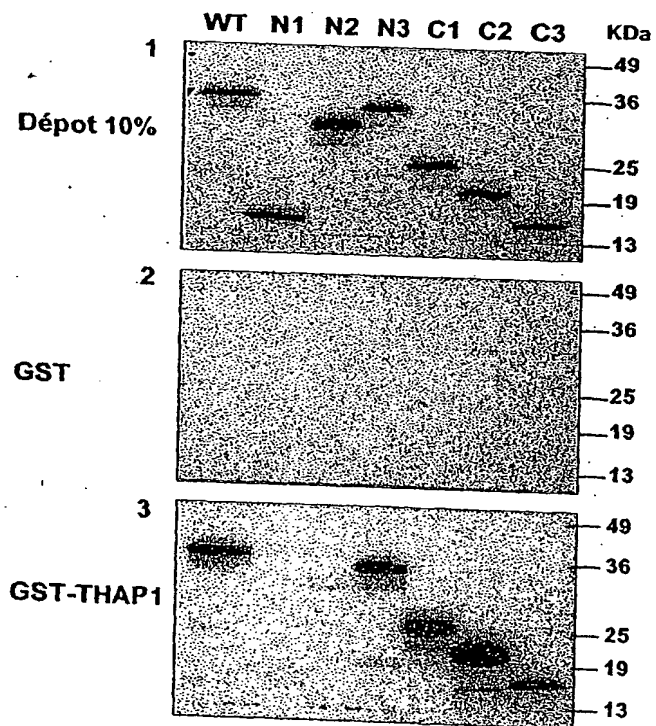


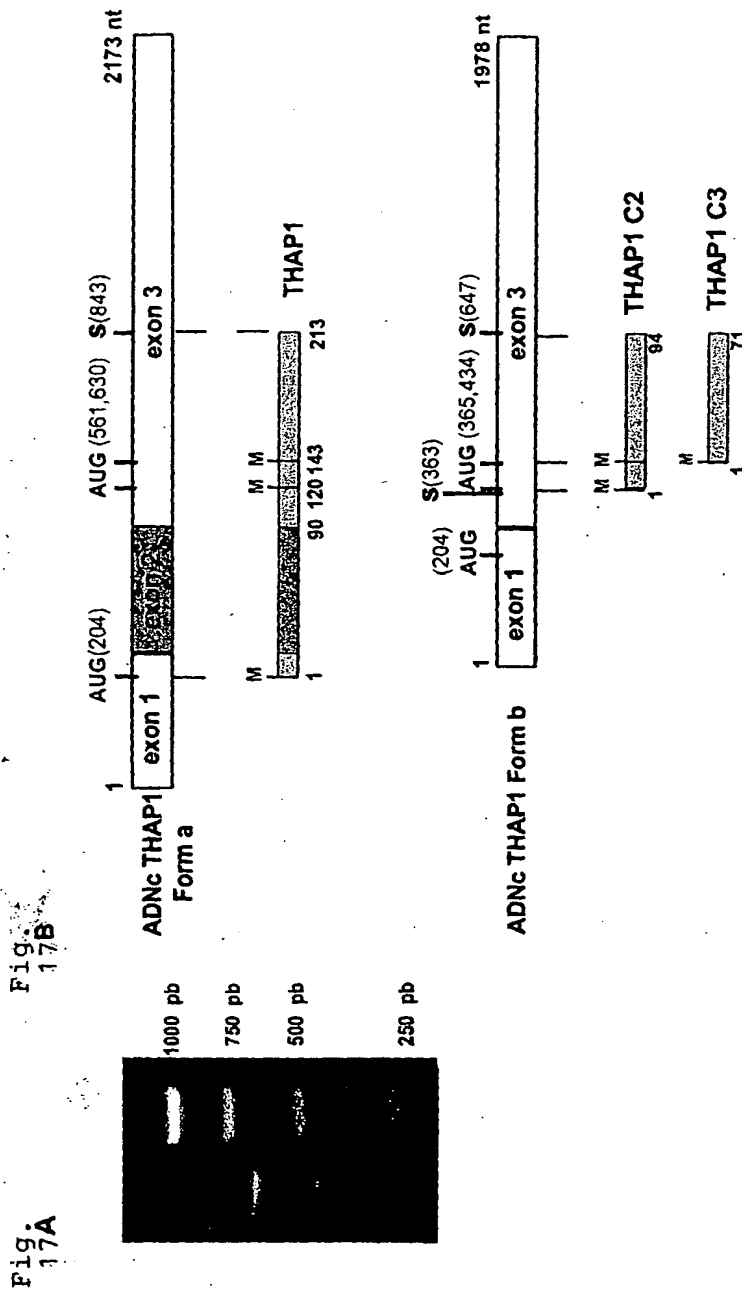
FIGURE 16

A



B





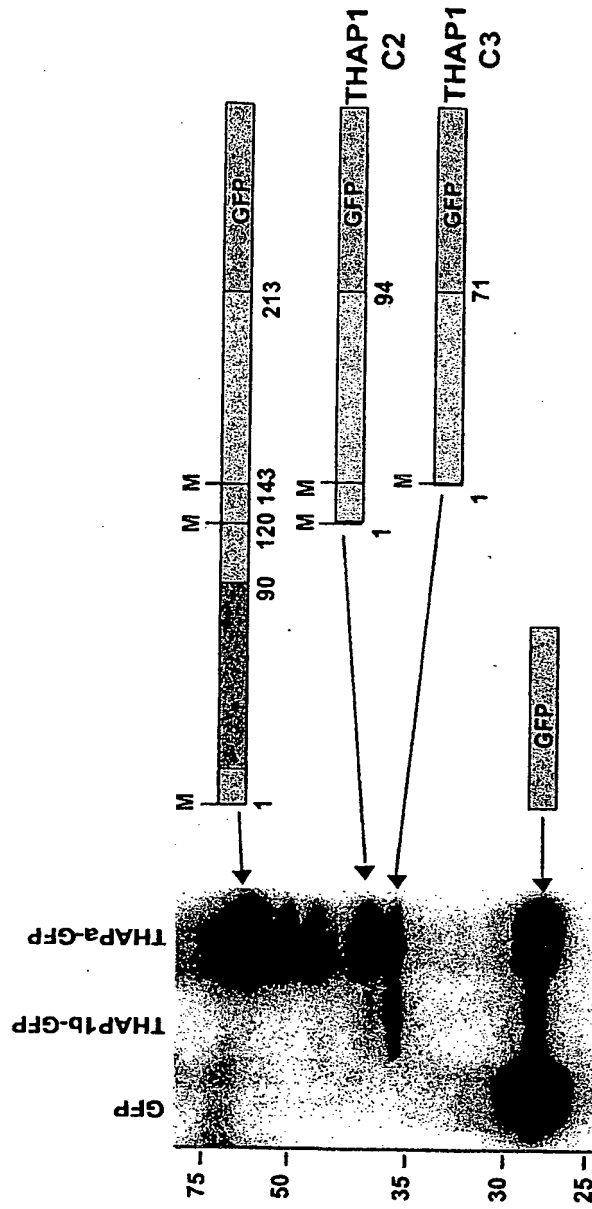


FIG. 17

Fig.  
17 C

Figure 18

GGGCAT ACTAC TGGCAA  
 GGGCAA ACTGT GGGCAT  
 GGGCAT ACTAC TGGCAA  
 GGGCAA ACTAC TGGCAA  
 GGGCCA GTTCG TTGCAA  
 GGGCAT GTAC TGGCAA  
 GGGCAA CTGT GGGCAA  
 GGGCAA CACTAC TGGCAA  
 GGGCAA AGTAC TGGCAA

A

1) DR-5 Consensus Motif  
 GGGCAAAnnnnnTGGCAA  
 (DR-4, DR-6)

TTGCCA GTACTAAGTGT GGGCAA  
 CTGCCA GTACATAGTGT GGGCAA  
 TTGCCA GTACTAAGTGT GGGCAA  
 CTGCCA GTAGATACTGT GGGCAA  
 TTGCCA GTAGTTAGGTGT GGGCGA  
 TTGCCA GTAGTTAGTGT GGGCAA  
 TTGCCA GTACCTACTAA GGGCAA  
 TTGCCA GTAGTTAGTGT GGGCAG  
 CTGCCA GTAGTAAGTGT GGGCAG

B

2) ER-11 Consensus Motif  
 TTGCCAAnnnnnnnnnnnGGGCAA  
 (ER-12)

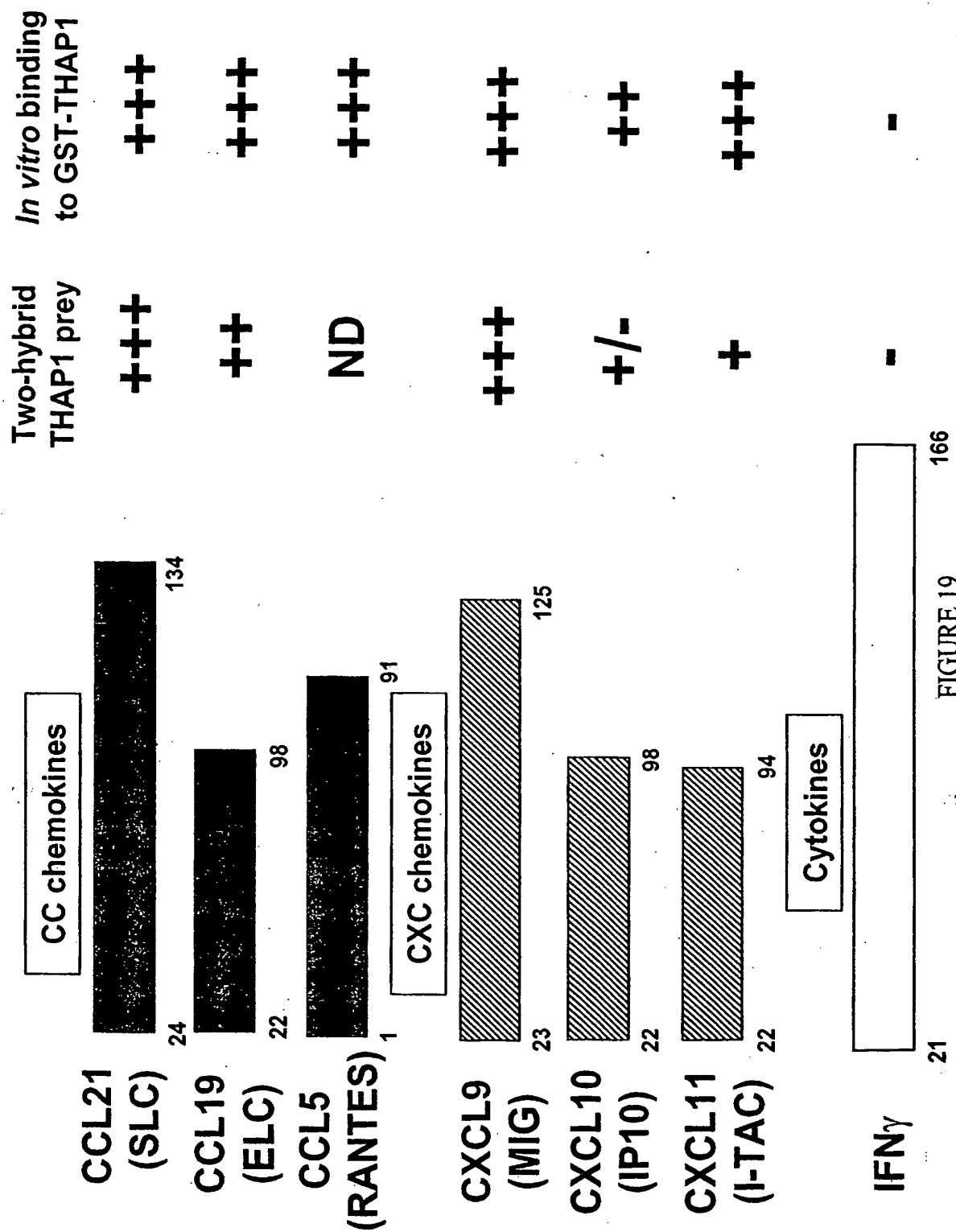
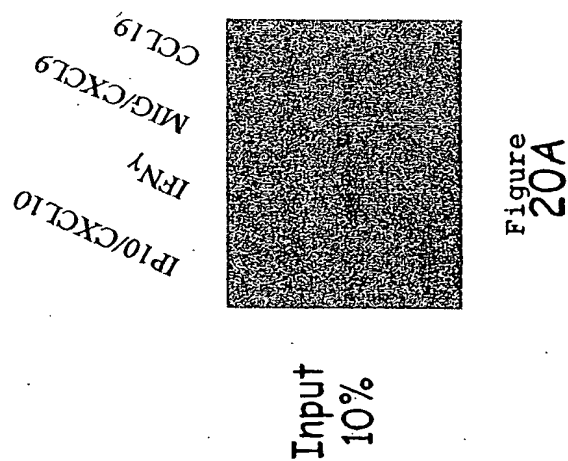
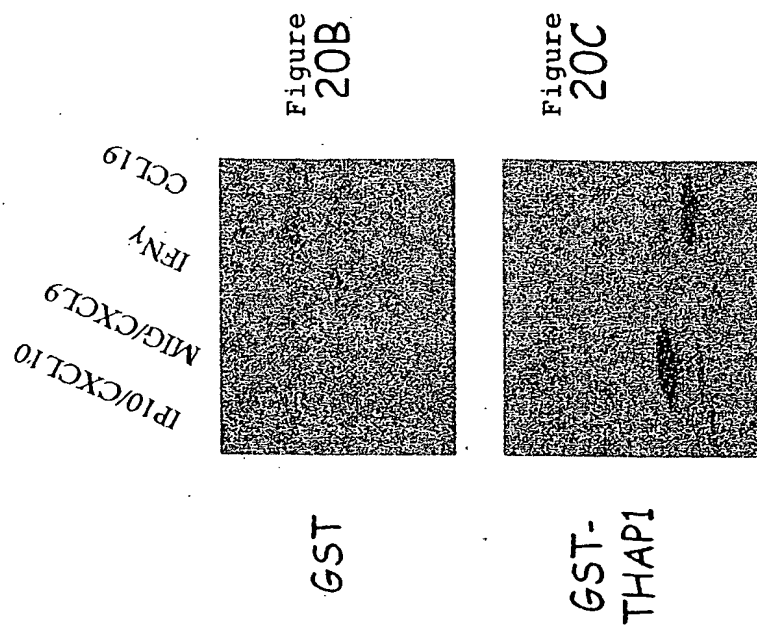


FIGURE 19





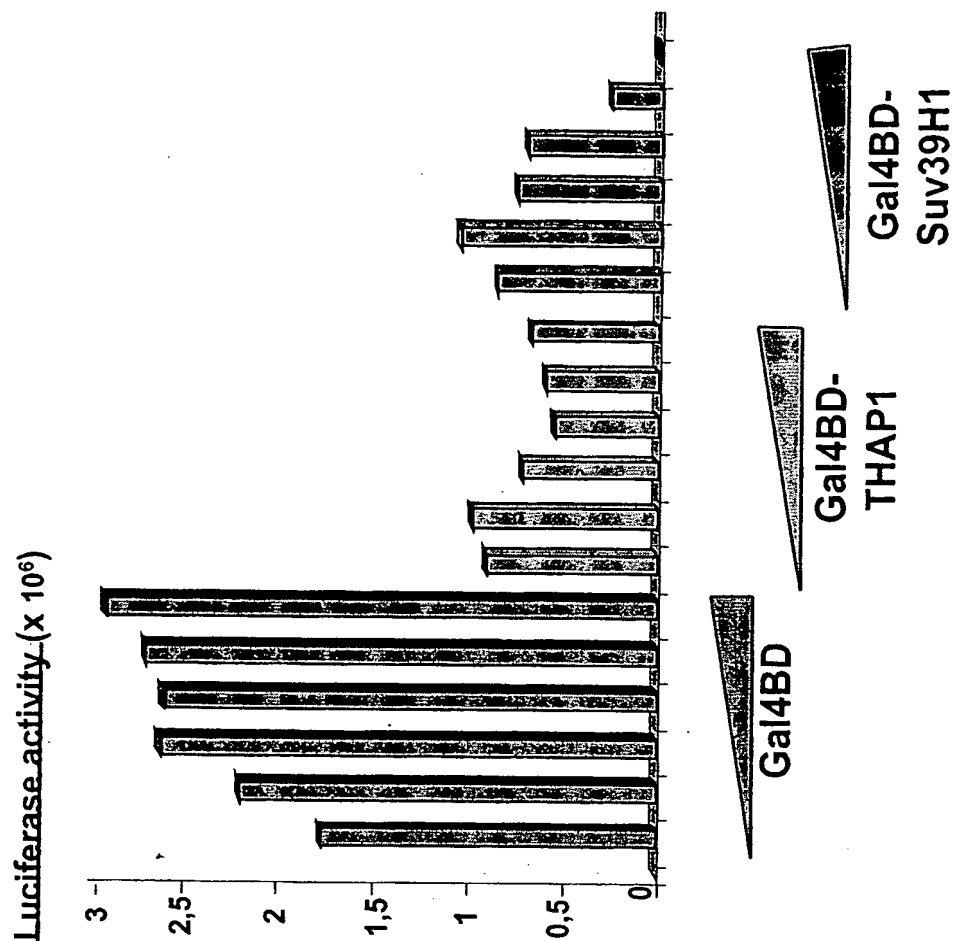


Fig. 21B

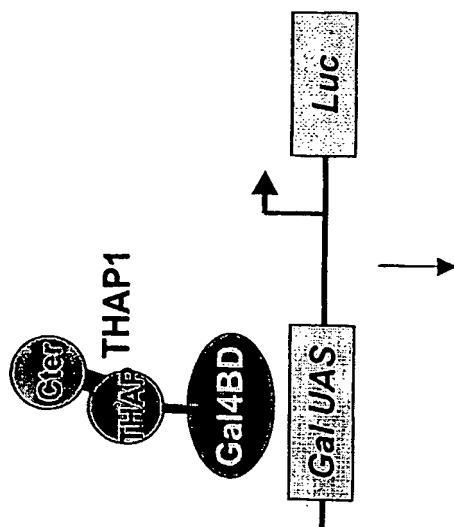


Fig. 21A

Figure 22A

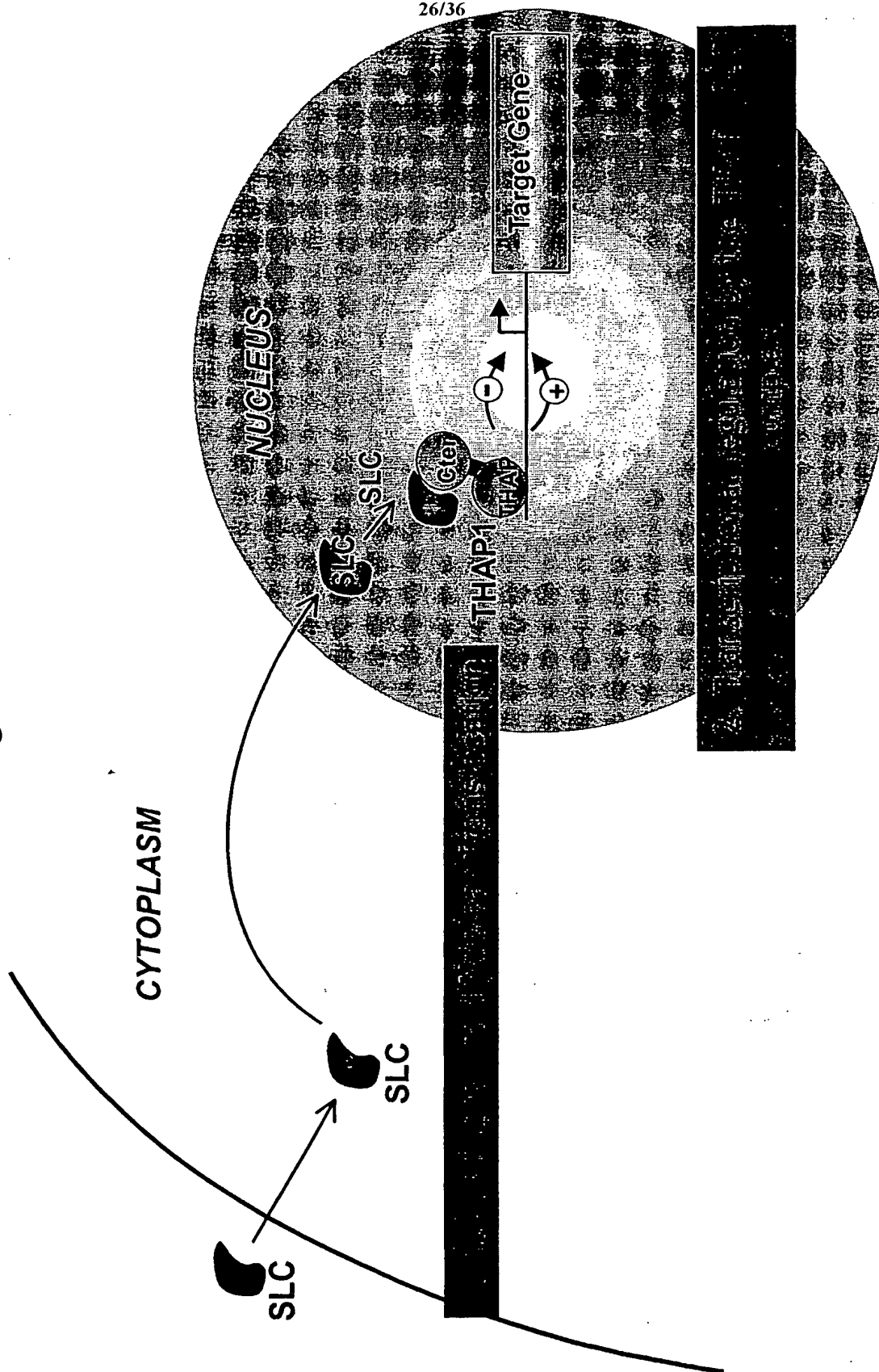


Figure 22B

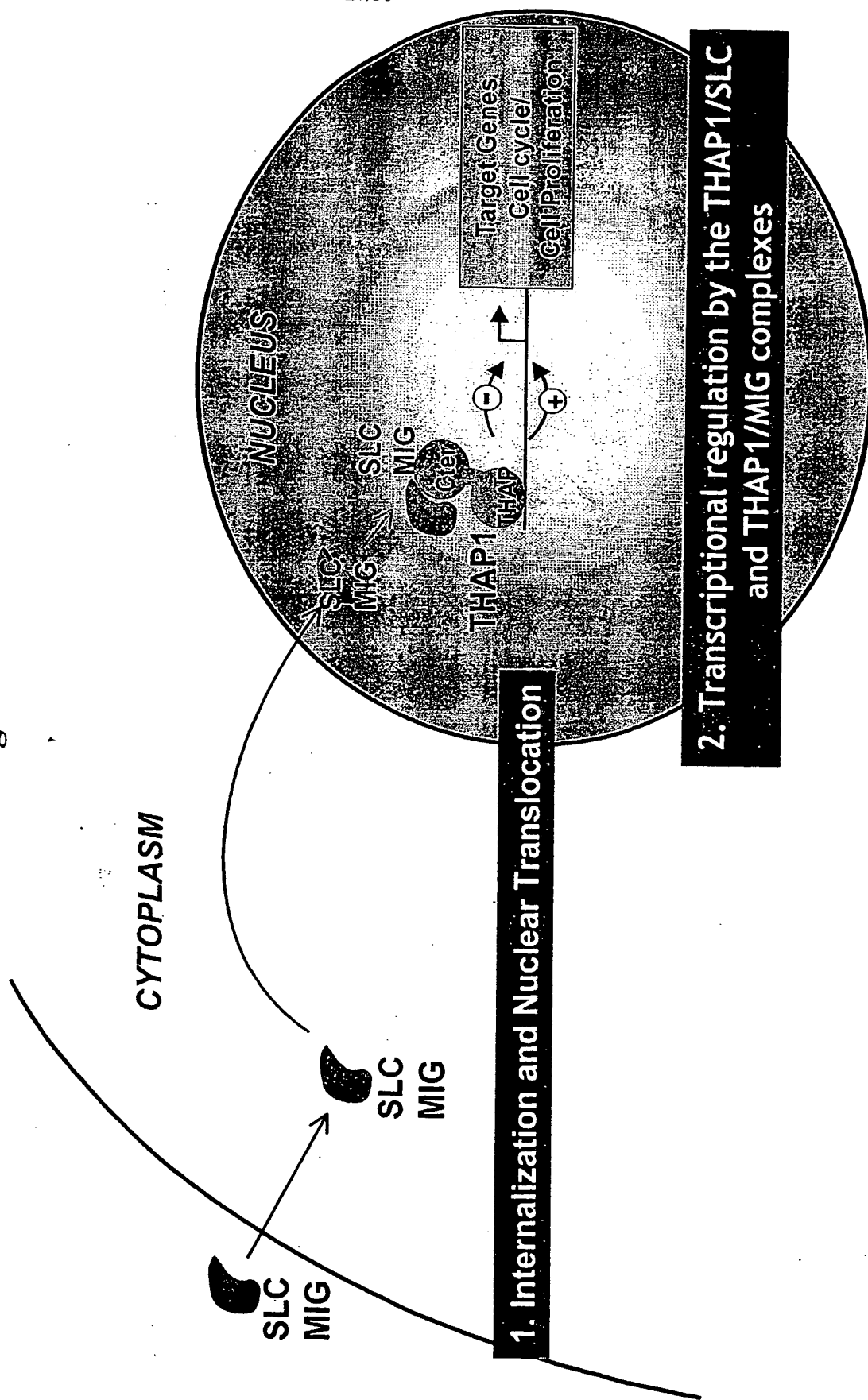
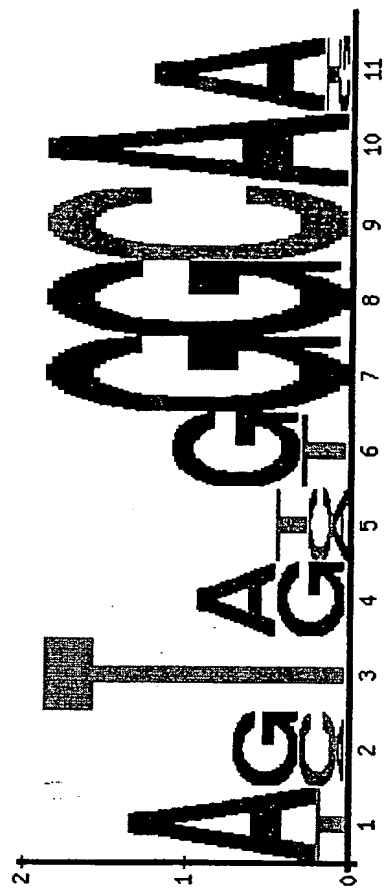


Figure 23

661 cttctctgcc. ctgggccaag cctgcccag cctctctgtc ctctgcctgc ccagctggac  
721 atctctgggc ctctctggag accagtgggg tgggctgtgg gggcgtcata ttgcccctggc  
781 ttggcatccc tcttgtggct gtacccctcc cagcagcccc aggactagca agtccccgag  
841 atgggggtgg ggacagtggg tgatgcaaa ggttgtgggg gcaggggcgg ggcaggagca  
901 ggaagggtccc ctgagttccc tcacctggg cagagataaa aggagcacag ttccaggcgg  
961 ggctgagcta gggcgtagct gtgatttcag gggcacctct ggcggctgcc gtgatttgag  
1021 aatctcgggt ctcttggctg actgacctg ggagactgtg gatgaataat gctggtgagt

The human *Fucosyltransferase TVII* promoter

Figure 24



Consensus sequence of THAP1-Responsive Element (THRE)

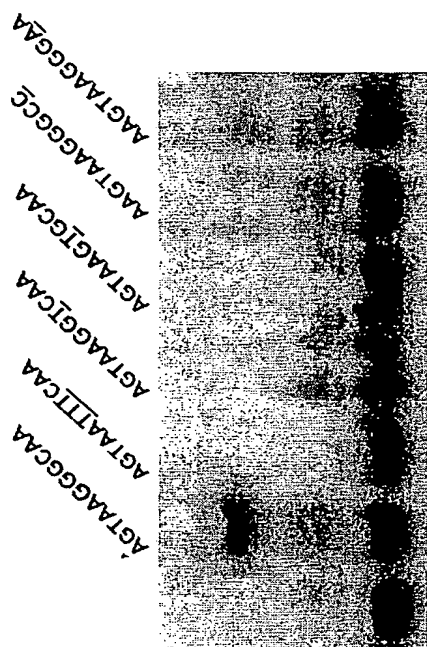


Figure 25A

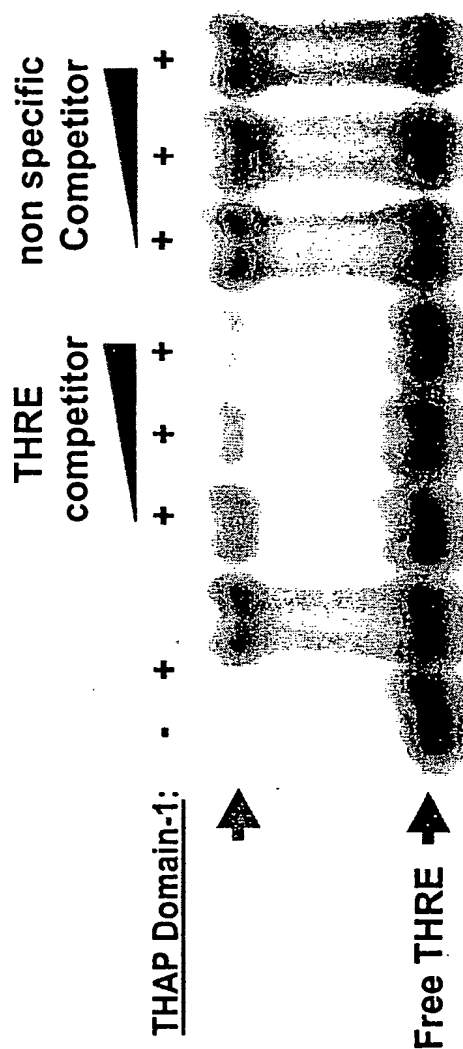


Figure 25B

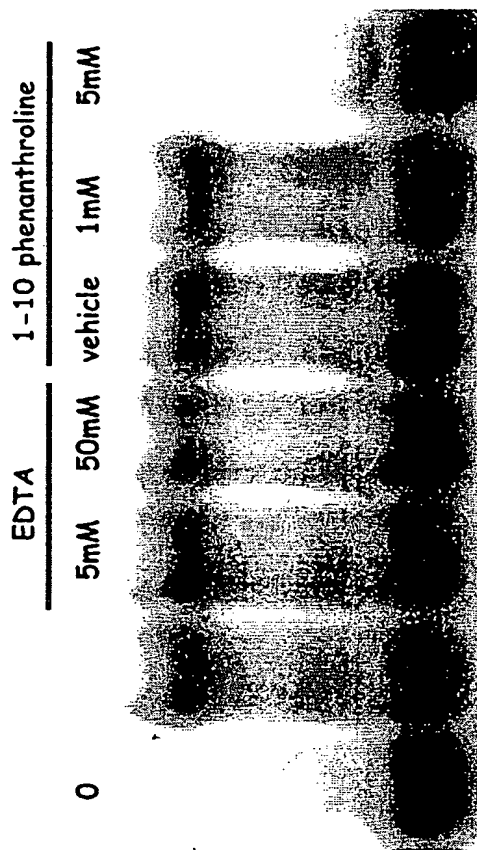


Figure 26A

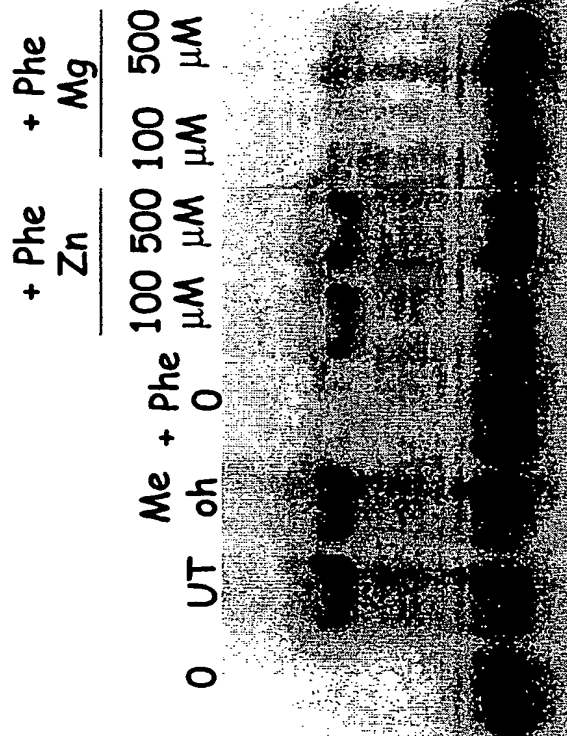
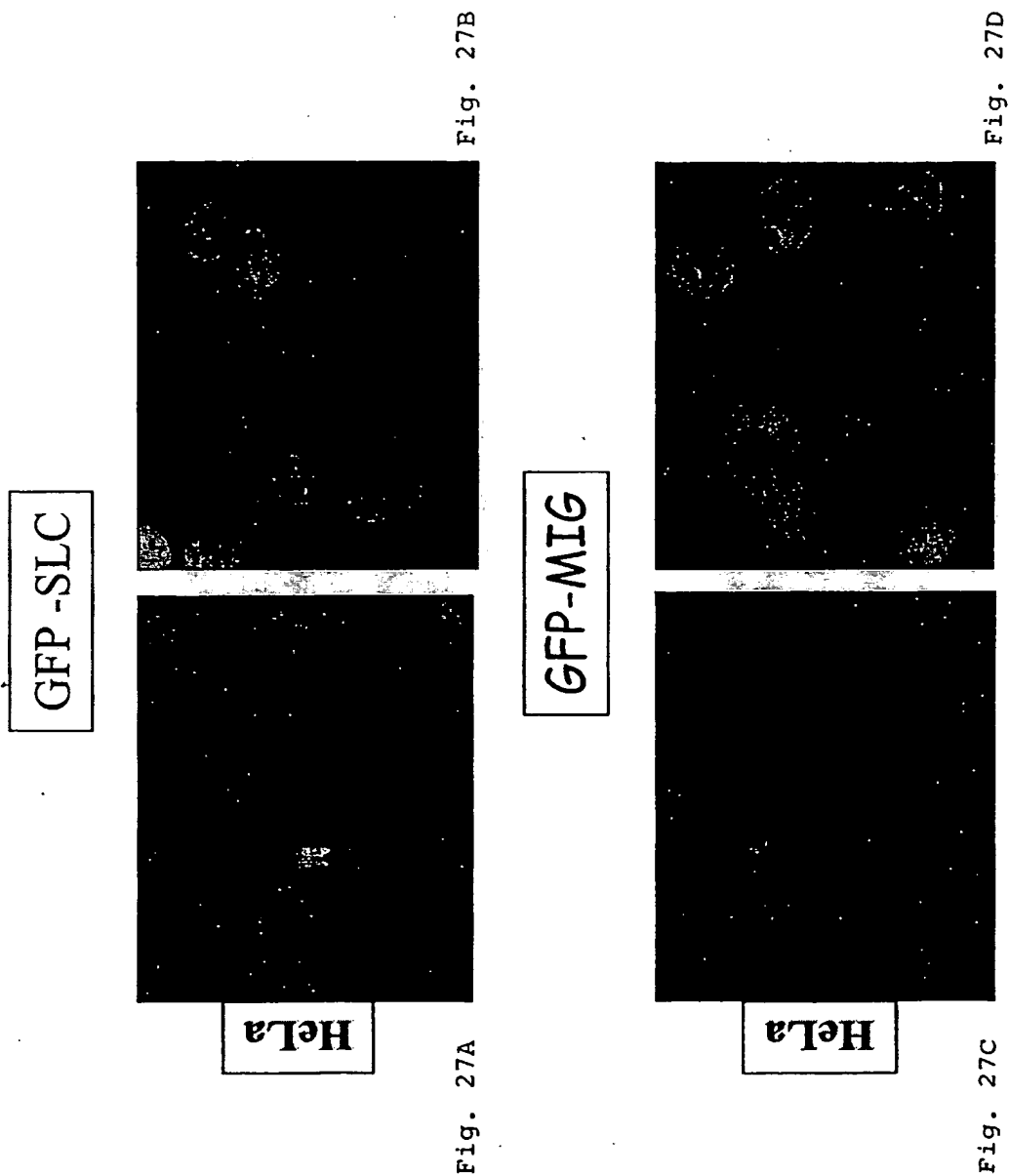


Figure 26B



# Figure 27



# Figure 28

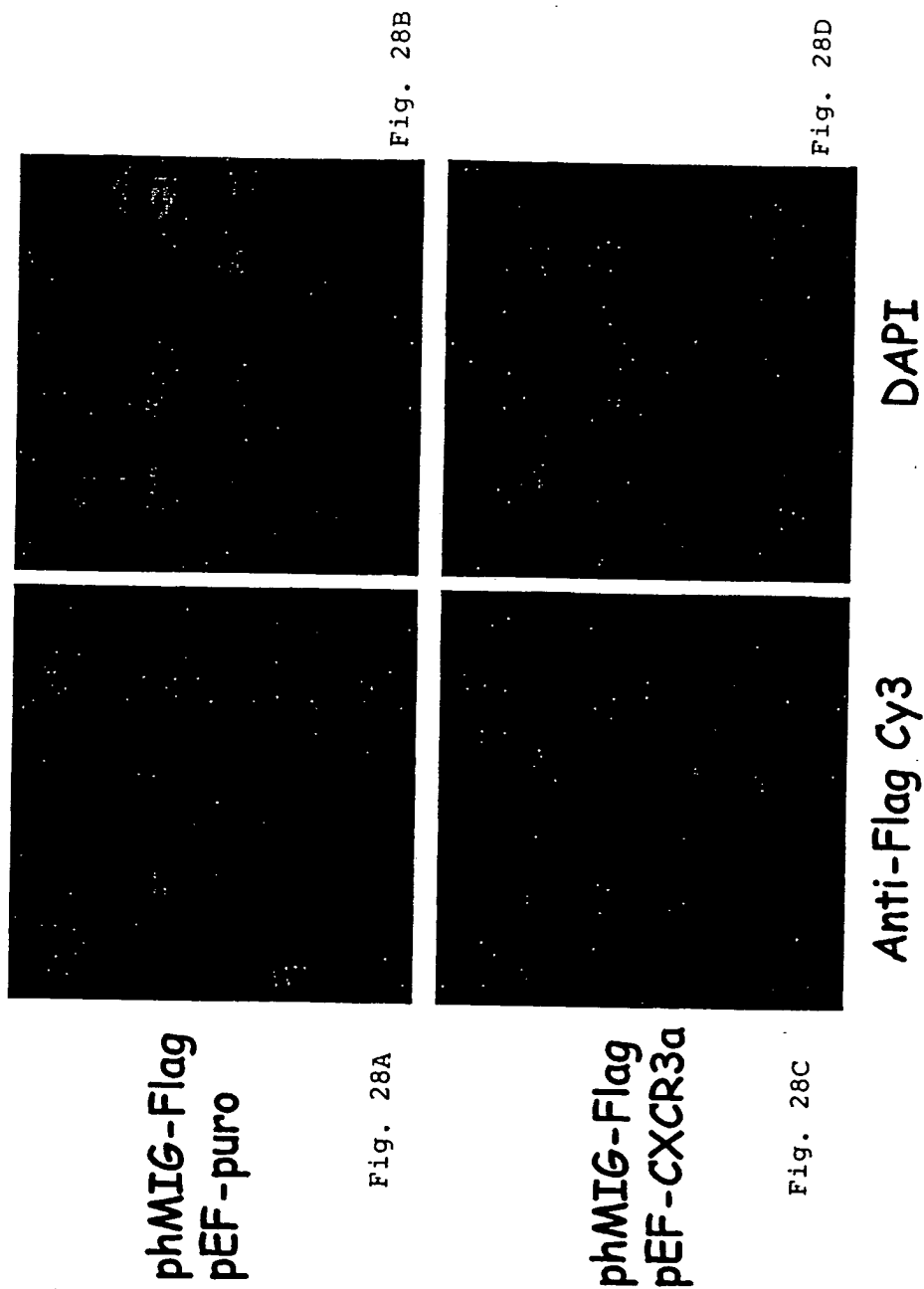




Fig. 29A

Fig. 29B

Fig. 29C

U2OS: phMIG-flag + pEF-CXCR3a

## Figure 30

### The human *Survivin* promoter

Hs Survivin/BIRC5 promoter fragment (GenBank NT 010641.14|Hs17 10798 Homo sapiens chromosome 17 genomic contig; nucleotides 10102350-10102668)

35/36

```

1  cgtcgctggg tgcaccgga ccacgggcag agccacgcgg cgggaggact acaactcccg
61  gcacaccccg cgccgccccg cctctactcc cagaaggccg cggggggtgg accgcctaag
121 agggcgtgcg ctccgacat gcccgcggc gcgccattaa ccgccagatt tgaatcgcg
181 gaccggttg cagagggtgc ggcgggcgca tgggtgccc gacgtgccc cctgcctggc
241 agccctttct caaggaccac cgcattctta cattcaagaa ctggcccttc ttggagggct
301 gcgcctgcac ccggagcg

```

# Figure 31

## The human *Ubiquitin specific protease 16* promoter

Hs USP16 promoter ; range -499 to 100 >EP73421 (EPD database at <http://www.epd.isb-sib.ch/>)

TGGTGCCGTGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGAAAGATGCACTCCAGCCTGGGCGACAGAGGATATT  
 CCGTCTCAAAACAACAAAAATCACTCGCTGCGTTTTTATTCTGACATGGTGCAGGAAGGTAAATTCAGACAACCTTAGGTAC  
 TCAGTTTTAGAAGTCGACAGGACAGAAATTACGGAAACAAATTTAAGCGTTCCCGCTTTAGCTCCAAATAAATGTGTTCCAGAA  
 AGGTAACCATCTAGGAACTCCAAGGCTCAGACCACCGGATGCCCACACITCAGGAGCATTTATATACTTCGTGGTTAT  
 GTCAGAGACGAGAAACCCATTGACAACCAACCCCTAAACCGCAACATCCGGCGCAAGCCGACGCGCGCAGATTTACTA  
 GCGTCAGAGCCGATGGTCCCGGAGGTGGGGTGGTGGCTAGCCACTTCCCAATAATGCCGCTTCGGGAAGT  
 TATTGCTTCCAGGGGTCACITCGGCTTCGACICCGICICATTCGICACCCAGGAGGAGACGCTGGCTGCCCAGC  
 CCAAAGGCCCATGAGGGGA

## SEQUENCE LISTING

<110> ENDOCUBE SAS  
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE - CNRS  
GIRARD, Jean-Philippe  
AMALRIC, Francois  
ROUSSIGNE, Myriam  
CLOUAIRE, Thomas

<120> THAP PROTEINS AS NUCLEAR RECEPTORS FOR  
CHEMOKINES AND ROLES IN TRANSCRIPTIONAL REGULATION, CELL  
PROLIFERATION AND CELL DIFFERENTIATION

<130> BIOBANK.012vpc

<150> US 60/432699

<151> 2002-12-10

<150> US 60/485027

<151> 2003-07-03

<160> 535

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 74

<212> PRT

<213> Artificial Sequence

<220>

<223> THAP domain consensus

<221> UNSURE

<222> 2-5, 7-21, 23-31, 33-49, 51-52, 55-73

<223> Xaa = any of the twenty amino acids

<400> 1

Cys	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa
1				5					10					15		
Xaa	Xaa	Xaa	Xaa	Xaa	Pro	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Trp	
			20					25				30				
Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa
		35					40				45					
Xaa	Cys	Xaa	Xaa	His	Phe	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa
	50				55						60					
Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Pro							
65					70											

<210> 2

<211> 81

<212> PRT

<213> Artificial Sequence

<220>

<223> THAP domain consensus

&lt;221&gt; UNSURE

&lt;222&gt; 3-4, 6-9, 11-21, 24, 27-35, 37-41, 43-53, 56, 59-62, 64-71, 74-75, 80

&lt;223&gt; Xaa = any of the twenty amino acids

&lt;400&gt; 2

```

Met Pro Xaa Xaa Cys Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa
1      5      10      15
Xaa Xaa Xaa Xaa Xaa Phe His Xaa Phe Pro Xaa Xaa Xaa Xaa Xaa
20      25      30
Xaa Xaa Xaa Trp Xaa Xaa Xaa Xaa Xaa Arg Xaa Xaa Xaa Xaa Xaa
35      40      45
Xaa Xaa Xaa Xaa Xaa Cys Ser Xaa His Phe Xaa Xaa Xaa Phe Xaa
50      55      60
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Leu Lys Xaa Xaa Ala Val Pro Thr Xaa
65      70      75      80
Phe

```

&lt;210&gt; 3

&lt;211&gt; 213

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 3

```

Met Val Gln Ser Cys Ser Ala Tyr Gly Cys Lys Asn Arg Tyr Asp Lys
1      5      10      15
Asp Lys Pro Val Ser Phe His Lys Phe Pro Leu Thr Arg Pro Ser Leu
20      25      30
Cys Lys Glu Trp Glu Ala Ala Val Arg Arg Lys Asn Phe Lys Pro Thr
35      40      45
Lys Tyr Ser Ser Ile Cys Ser Glu His Phe Thr Pro Asp Cys Phe Lys
50      55      60
Arg Glu Cys Asn Asn Lys Leu Leu Lys Glu Asn Ala Val Pro Thr Ile
65      70      75      80
Phe Leu Cys Thr Glu Pro His Asp Lys Lys Glu Asp Leu Leu Glu Pro
85      90      95
Gln Glu Gln Leu Pro Pro Pro Pro Leu Pro Pro Pro Val Ser Gln Val
100      105      110
Asp Ala Ala Ile Gly Leu Leu Met Pro Pro Leu Gln Thr Pro Val Asn
115      120      125
Leu Ser Val Phe Cys Asp His Asn Tyr Thr Val Glu Asp Thr Met His
130      135      140
Gln Arg Lys Arg Ile His Gln Leu Glu Gln Gln Val Glu Lys Leu Arg
145      150      155      160
Lys Lys Leu Lys Thr Ala Gln Gln Arg Cys Arg Arg Gln Glu Arg Gln
165      170      175
Leu Glu Lys Leu Lys Glu Val Val His Phe Gln Lys Glu Lys Asp Asp
180      185      190
Val Ser Glu Arg Gly Tyr Val Ile Leu Pro Asn Asp Tyr Phe Glu Ile
195      200      205
Val Glu Val Pro Ala
210

```

&lt;210&gt; 4

&lt;211&gt; 228

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 4

```

Met Pro Thr Asn Cys Ala Ala Ala Gly Cys Ala Thr Thr Tyr Asn Lys
1          5          10          15
His Ile Asn Ile Ser Phe His Arg Phe Pro Leu Asp Pro Lys Arg Arg
20          25          30
Lys Glu Trp Val Arg Leu Val Arg Arg Lys Asn Phe Val Pro Gly Lys
35          40          45
His Thr Phe Leu Cys Ser Lys His Phe Glu Ala Ser Cys Phe Asp Leu
50          55          60
Thr Gly Gln Thr Arg Arg Leu Lys Met Asp Ala Val Pro Thr Ile Phe
65          70          75          80
Asp Phe Cys Thr His Ile Lys Ser Met Lys Leu Lys Ser Arg Asn Leu
85          90          95
Leu Lys Lys Asn Asn Ser Cys Ser Pro Ala Gly Pro Ser Asn Leu Lys
100         105         110
Ser Asn Ile Ser Ser Gln Gln Val Leu Leu Glu His Ser Tyr Ala Phe
115         120         125
Arg Asn Pro Met Glu Ala Lys Lys Arg Ile Ile Lys Leu Glu Lys Glu
130         135         140
Ile Ala Ser Leu Arg Arg Lys Met Lys Thr Cys Leu Gln Lys Glu Arg
145         150         155         160
Arg Ala Thr Arg Arg Trp Ile Lys Ala Thr Cys Leu Val Lys Asn Leu
165         170         175
Glu Ala Asn Ser Val Leu Pro Lys Gly Thr Ser Glu His Met Leu Pro
180         185         190
Thr Ala Leu Ser Ser Leu Pro Leu Glu Asp Phe Lys Ile Leu Glu Gln
195         200         205
Asp Gln Gln Asp Lys Thr Leu Leu Ser Leu Asn Leu Lys Gln Thr Lys
210         215         220
Ser Thr Phe Ile
225

```

&lt;210&gt; 5

&lt;211&gt; 239

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 5

```

Met Pro Lys Ser Cys Ala Ala Arg Gln Cys Cys Asn Arg Tyr Ser Ser
1          5          10          15
Arg Arg Lys Gln Leu Thr Phe His Arg Phe Pro Phe Ser Arg Pro Glu
20          25          30
Leu Leu Lys Glu Trp Val Leu Asn Ile Gly Arg Gly Asn Phe Lys Pro
35          40          45
Lys Gln His Thr Val Ile Cys Ser Glu His Phe Arg Pro Glu Cys Phe
50          55          60
Ser Ala Phe Gly Asn Arg Lys Asn Leu Lys His Asn Ala Val Pro Thr
65          70          75          80
Val Phe Ala Phe Gln Asp Pro Thr Gln Gln Val Arg Glu Asn Thr Asp
85          90          95
Pro Ala Ser Glu Arg Gly Asn Ala Ser Ser Ser Gln Lys Glu Lys Val
100         105         110
Leu Pro Glu Ala Gly Ala Gly Glu Asp Ser Pro Gly Arg Asn Met Asp
115         120         125
Thr Ala Leu Glu Glu Leu Gln Leu Pro Pro Asn Ala Glu Gly His Val
130         135         140
Lys Gln Val Ser Pro Arg Arg Pro Gln Ala Thr Glu Ala Val Gly Arg
145         150         155         160
Pro Thr Gly Pro Ala Gly Leu Arg Arg Thr Pro Asn Lys Gln Pro Ser

```



```
<210> 6
<211> 577
<212> PRT
<213> Homo sapiens
```

-4-

Arg Gln Asn Lys Ser Gln Val Cys Cys Leu Arg Glu Gln Val Glu Lys  
 355 360 365  
 Lys Asn Gly Glu Leu Lys Ser Leu Arg Gln Arg Val Ser Arg Ser Asp  
 370 375 380  
 Ser Gln Val Arg Lys Leu Gln Glu Lys Leu Asp Glu Leu Arg Arg Val  
 385 390 395 400  
 Ser Val Pro Tyr Pro Ser Ser Leu Leu Ser Pro Ser Arg Glu Pro Pro  
 405 410 415  
 Lys Met Asn Pro Val Val Glu Pro Leu Ser Trp Met Leu Gly Thr Trp  
 420 425 430  
 Leu Ser Asp Pro Pro Gly Ala Gly Thr Tyr Pro Thr Leu Gln Pro Phe  
 435 440 445  
 Gln Tyr Leu Glu Glu Val His Ile Ser His Val Gly Gln Pro Met Leu  
 450 455 460  
 Asn Phe Ser Phe Asn Ser Phe His Pro Asp Thr Arg Lys Pro Met His  
 465 470 475 480  
 Arg Glu Cys Gly Phe Ile Arg Leu Lys Pro Asp Thr Asn Lys Val Ala  
 485 490 495  
 Phe Val Ser Ala Gln Asn Thr Gly Val Val Glu Val Glu Glu Gly Glu  
 500 505 510  
 Val Asn Gly Gln Glu Leu Cys Ile Ala Ser His Ser Ile Ala Arg Ile  
 515 520 525  
 Ser Phe Ala Lys Glu Pro His Val Glu Gln Ile Thr Arg Lys Phe Arg  
 530 535 540  
 Leu Asn Ser Glu Gly Lys Leu Glu Gln Thr Val Ser Met Ala Thr Thr  
 545 550 555 560  
 Thr Gln Pro Met Thr Gln His Leu His Val Thr Tyr Lys Lys Val Thr  
 565 570 575  
 Pro

<210> 7  
 <211> 395  
 <212> PRT  
 <213> Homo sapiens

<400> 7  
 Met Pro Arg Tyr Cys Ala Ala Ile Cys Cys Lys Asn Arg Arg Gly Arg  
 1 5 10 15  
 Asn Asn Lys Asp Arg Lys Leu Ser Phe Tyr Pro Phe Pro Leu His Asp  
 20 25 30  
 Lys Glu Arg Leu Glu Lys Trp Leu Lys Asn Met Lys Arg Asp Ser Trp  
 35 40 45  
 Val Pro Ser Lys Tyr Gln Phe Leu Cys Ser Asp His Phe Thr Pro Asp  
 50 55 60  
 Ser Leu Asp Ile Arg Trp Gly Ile Arg Tyr Leu Lys Gln Thr Ala Val  
 65 70 75 80  
 Pro Thr Ile Phe Ser Leu Pro Glu Asp Asn Gln Gly Lys Asp Pro Ser  
 85 90 95  
 Lys Lys Lys Ser Gln Lys Lys Asn Leu Glu Asp Glu Lys Glu Val Cys  
 100 105 110  
 Pro Lys Ala Lys Ser Glu Glu Ser Phe Val Leu Asn Glu Thr Lys Lys  
 115 120 125  
 Asn Ile Val Asn Thr Asp Val Pro His Gln His Pro Glu Leu Leu His  
 130 135 140  
 Ser Ser Ser Leu Val Lys Pro Pro Ala Pro Lys Thr Gly Ser Ile Gln  
 145 150 155 160  
 Asn Asn Met Leu Thr Leu Asn Leu Val Lys Gln His Thr Gly Lys Pro  
 165 170 175  
 Glu Ser Thr Leu Glu Thr Ser Val Asn Gln Asp Thr Gly Arg Gly Gly

```

      180      185      190
Phe His Thr Cys Phe Glu Asn Leu Asn Ser Thr Thr Ile Thr Leu Thr
      195      200      205
Thr Ser Asn Ser Glu Ser Ile His Gln Ser Leu Glu Thr Gln Glu Val
      210      215      220
Leu Glu Val Thr Thr Ser His Leu Ala Asn Pro Asn Phe Thr Ser Asn
225      230      235      240
Ser Met Glu Ile Lys Ser Ala Gln Glu Asn Pro Phe Leu Phe Ser Thr
      245      250      255
Ile Asn Gln Thr Val Glu Glu Leu Asn Thr Asn Lys Glu Ser Val Ile
      260      265      270
Ala Ile Phe Val Pro Ala Glu Asn Ser Lys Pro Ser Val Asn Ser Phe
      275      280      285
Ile Ser Ala Gln Lys Glu Thr Thr Glu Met Glu Asp Thr Asp Ile Glu
      290      295      300
Asp Ser Leu Tyr Lys Asp Val Asp Tyr Gly Thr Glu Val Leu Gln Ile
305      310      315      320
Glu His Ser Tyr Cys Arg Gln Asp Ile Asn Lys Glu His Leu Trp Gln
      325      330      335
Lys Val Ser Lys Leu His Ser Lys Ile Thr Leu Leu Glu Leu Lys Glu
      340      345      350
Gln Gln Thr Leu Gly Arg Leu Lys Ser Leu Glu Ala Leu Ile Arg Gln
      355      360      365
Leu Lys Gln Glu Asn Trp Leu Ser Glu Glu Asn Val Lys Ile Ile Glu
      370      375      380
Asn His Phe Thr Thr Tyr Glu Val Thr Met Ile
385      390      395

```

<210> 8  
 <211> 222  
 <212> PRT  
 <213> Homo sapiens

```

<400> 8
Met Val Lys Cys Cys Ser Ala Ile Gly Cys Ala Ser Arg Cys Leu Pro
  1      5      10      15
Asn Ser Lys Leu Lys Gly Leu Thr Phe His Val Phe Pro Thr Asp Glu
      20      25      30
Asn Ile Lys Arg Lys Trp Val Leu Ala Met Lys Arg Leu Asp Val Asn
      35      40      45
Ala Ala Gly Ile Trp Glu Pro Lys Lys Gly Asp Val Leu Cys Ser Arg
      50      55      60
His Phe Lys Lys Thr Asp Phe Asp Arg Ser Ala Pro Asn Ile Lys Leu
      65      70      75      80
Lys Pro Gly Val Ile Pro Ser Ile Phe Asp Ser Pro Tyr His Leu Gln
      85      90      95
Gly Lys Arg Glu Lys Leu His Cys Arg Lys Asn Phe Thr Leu Lys Thr
      100      105      110
Val Pro Ala Thr Asn Tyr Asn His His Leu Val Gly Ala Ser Ser Cys
      115      120      125
Ile Glu Glu Phe Gln Ser Gln Phe Ile Phe Glu His Ser Tyr Ser Val
      130      135      140
Met Asp Ser Pro Lys Lys Leu Lys His Lys Leu Asp His Val Ile Gly
      145      150      155      160
Glu Leu Glu Asp Thr Lys Glu Ser Leu Arg Asn Val Leu Asp Arg Glu
      165      170      175
Lys Arg Phe Gln Lys Ser Leu Arg Lys Thr Ile Arg Glu Leu Lys Asp
      180      185      190
Glu Cys Leu Ile Ser Gln Glu Thr Ala Asn Arg Leu Asp Thr Phe Cys
      195      200      205

```

Trp Asp Cys Cys Gln Glu Ser Ile Glu Gln Asp Tyr Ile Ser  
 210 215 220

<210> 9  
 <211> 309  
 <212> PRT  
 <213> Homo sapiens

<400> 9  
 Met Pro Arg His Cys Ser Ala Ala Gly Cys Cys Thr Arg Asp Thr Arg  
 1 5 10 15  
 Glu Thr Arg Asn Arg Gly Ile Ser Phe His Arg Leu Pro Lys Lys Asp  
 20 25 30  
 Asn Pro Arg Arg Gly Leu Trp Leu Ala Asn Cys Gln Arg Leu Asp Pro  
 35 40 45  
 Ser Gly Gln Gly Leu Trp Asp Pro Ala Ser Glu Tyr Ile Tyr Phe Cys  
 50 55 60  
 Ser Lys His Phe Glu Glu Asp Cys Phe Glu Leu Val Gly Ile Ser Gly  
 65 70 75 80  
 Tyr His Arg Leu Lys Glu Gly Ala Val Pro Thr Ile Phe Glu Ser Phe  
 85 90 95  
 Ser Lys Leu Arg Arg Thr Thr Lys Thr Lys Gly His Ser Tyr Pro Pro  
 100 105 110  
 Gly Pro Pro Glu Val Ser Arg Leu Arg Arg Cys Arg Lys Arg Cys Ser  
 115 120 125  
 Glu Gly Arg Gly Pro Thr Thr Pro Phe Ser Pro Pro Pro Pro Ala Asp  
 130 135 140  
 Val Thr Cys Phe Pro Val Glu Glu Ala Ser Ala Pro Ala Thr Leu Pro  
 145 150 155 160  
 Ala Ser Pro Ala Gly Arg Leu Glu Pro Gly Leu Ser Ser Pro Phe Ser  
 165 170 175  
 Asp Leu Leu Gly Pro Leu Gly Ala Gln Ala Asp Glu Ala Gly Cys Ser  
 180 185 190  
 Ala Gln Pro Ser Pro Glu Arg Gln Pro Ser Pro Leu Glu Pro Arg Pro  
 195 200 205  
 Val Ser Pro Ser Ala Tyr Met Leu Arg Leu Pro Pro Pro Ala Gly Ala  
 210 215 220  
 Tyr Ile Gln Asn Glu His Ser Tyr Gln Val Gly Ser Ala Leu Leu Trp  
 225 230 235 240  
 Lys Arg Arg Ala Glu Ala Ala Leu Asp Ala Leu Asp Lys Ala Gln Arg  
 245 250 255  
 Gln Leu Gln Ala Cys Lys Arg Arg Glu Gln Arg Leu Arg Leu Arg Leu  
 260 265 270  
 Thr Lys Leu Gln Gln Glu Arg Ala Arg Glu Lys Arg Ala Gln Ala Asp  
 275 280 285  
 Ala Arg Gln Thr Leu Lys Glu His Val Gln Asp Phe Ala Met Gln Leu  
 290 295 300  
 Ser Ser Ser Met Ala  
 305

<210> 10  
 <211> 274  
 <212> PRT  
 <213> Homo sapiens

<400> 10  
 Met Pro Lys Tyr Cys Arg Ala Pro Asn Cys Ser Asn Thr Ala Gly Arg  
 1 5 10 15  
 Leu Gly Ala Asp Asn Arg Pro Val Ser Phe Tyr Lys Phe Pro Leu Lys

										20											25											30
Asp	Gly	Pro	Arg	Leu	Gln	Ala	Trp	Leu	Gln	His	Met	Gly	Cys	Glu	His																	
										35											40											45
Trp	Val	Pro	Ser	Cys	His	Gln	His	Leu	Cys	Ser	Glu	His	Phe	Thr	Pro																	
										50											55											60
Ser	Cys	Phe	Gln	Trp	Arg	Trp	Gly	Val	Arg	Tyr	Leu	Arg	Pro	Asp	Ala																	
										65											70											75
Val	Pro	Ser	Ile	Phe	Ser	Arg	Gly	Pro	Pro	Ala	Lys	Ser	Gln	Arg	Arg																	
										85											90											95
Thr	Arg	Ser	Thr	Gln	Lys	Pro	Val	Ser	Pro	Pro	Pro	Pro	Leu	Gln	Lys																	
										100											105											110
Asn	Thr	Pro	Leu	Pro	Gln	Ser	Pro	Ala	Ile	Pro	Val	Ser	Gly	Pro	Val																	
										115											120											125
Arg	Leu	Val	Val	Leu	Gly	Pro	Thr	Ser	Gly	Ser	Pro	Lys	Thr	Val	Ala																	
										130											135											140
Thr	Met	Leu	Leu	Thr	Pro	Leu	Ala	Pro	Ala	Pro	Thr	Pro	Glu	Arg	Ser																	
										145											150											155
Gln	Pro	Glu	Val	Pro	Ala	Gln	Gln	Ala	Gln	Thr	Gly	Leu	Gly	Pro	Val																	
										165											170											175
Leu	Gly	Ala	Leu	Gln	Arg	Arg	Val	Arg	Arg	Leu	Gln	Arg	Cys	Gln	Glu																	
										180											185											190
Arg	His	Gln	Ala	Gln	Leu	Gln	Ala	Leu	Glu	Arg	Leu	Ala	Gln	Gln	Leu																	
										195											200											205
His	Gly	Glu	Ser	Leu	Leu	Ala	Arg	Ala	Arg	Arg	Gly	Leu	Gln	Arg	Leu																	
										210											215											220
Thr	Thr	Ala	Gln	Thr	Leu	Gly	Pro	Glu	Glu	Ser	Gln	Thr	Phe	Thr	Ile																	
										225											230											235
Ile	Cys	Gly	Gly	Pro	Asp	Ile	Ala	Met	Val	Leu	Ala	Gln	Asp	Pro	Ala																	
										245											250											255
Pro	Ala	Thr	Val	Asp	Ala	Lys	Pro	Glu	Leu	Leu	Asp	Thr	Arg	Ile	Pro																	
										260											265											270
Ser Ala																																

<210> 11  
 <211> 903  
 <212> PRT  
 <213> Homo sapiens

<400> 11

Met	Thr	Arg	Ser	Cys	Ser	Ala	Val	Gly	Cys	Ser	Thr	Arg	Asp	Thr	Val
1				5					10					15	
Leu	Ser	Arg	Glu	Arg	Gly	Leu	Ser	Phe	His	Gln	Phe	Pro	Thr	Asp	Thr
			20					25					30		
Ile	Gln	Arg	Ser	Lys	Trp	Ile	Arg	Ala	Val	Asn	Arg	Val	Asp	Pro	Arg
			35				40					45			
Ser	Lys	Lys	Ile	Trp	Ile	Pro	Gly	Pro	Gly	Ala	Ile	Leu	Cys	Ser	Lys
			50			55					60				
His	Phe	Gln	Glu	Ser	Asp	Phe	Glu	Ser	Tyr	Gly	Ile	Arg	Arg	Lys	Leu
65					70					75				80	
Lys	Lys	Gly	Ala	Val	Pro	Ser	Val	Ser	Leu	Tyr	Lys	Ile	Pro	Gln	Gly
			85						90					95	
Val	His	Leu	Lys	Gly	Lys	Ala	Arg	Gln	Lys	Ile	Leu	Lys	Gln	Pro	Leu
			100					105					110		
Pro	Asp	Asn	Ser	Gln	Glu	Val	Ala	Thr	Glu	Asp	His	Asn	Tyr	Ser	Leu
			115				120					125			
Lys	Thr	Pro	Leu	Thr	Ile	Gly	Ala	Glu	Lys	Leu	Ala	Glu	Val	Gln	Gln
			130				135					140			
Met	Leu	Gln	Val	Ser	Lys	Lys	Arg	Leu	Ile	Ser	Val	Lys	Asn	Tyr	Arg
145					150					155					160

Met Ile Lys Lys Arg Lys Gly Leu Arg Leu Ile Asp Ala Leu Val Glu  
 165 170 175  
 Glu Lys Leu Leu Ser Glu Glu Thr Glu Cys Leu Leu Arg Ala Gln Phe  
 180 185 190  
 Ser Asp Phe Lys Trp Glu Leu Tyr Asn Trp Arg Glu Thr Asp Glu Tyr  
 195 200 205  
 Ser Ala Glu Met Lys Gln Phe Ala Cys Thr Leu Tyr Leu Cys Ser Ser  
 210 215 220  
 Lys Val Tyr Asp Tyr Val Arg Lys Ile Leu Lys Leu Pro His Ser Ser  
 225 230 235  
 Ile Leu Arg Thr Trp Leu Ser Lys Cys Gln Pro Ser Pro Gly Phe Asn  
 245 250 255  
 Ser Asn Ile Phe Ser Phe Leu Gln Arg Arg Val Glu Asn Gly Asp Gln  
 260 265 270  
 Leu Tyr Gln Tyr Cys Ser Leu Leu Ile Lys Ser Ile Pro Leu Lys Gln  
 275 280 285  
 Gln Leu Gln Trp Asp Pro Ser Ser His Ser Leu Gln Gly Phe Met Asp  
 290 295 300  
 Phe Gly Leu Gly Lys Leu Asp Ala Asp Glu Thr Pro Leu Ala Ser Glu  
 305 310 315 320  
 Thr Val Leu Leu Met Ala Val Gly Ile Phe Gly His Trp Arg Thr Pro  
 325 330 335  
 Leu Gly Tyr Phe Phe Val Asn Arg Ala Ser Gly Tyr Leu Gln Ala Gln  
 340 345 350  
 Leu Leu Arg Leu Thr Ile Gly Lys Leu Ser Asp Ile Gly Ile Thr Val  
 355 360 365  
 Leu Ala Val Thr Ser Asp Ala Thr Ala His Ser Val Gln Met Ala Lys  
 370 375 380  
 Ala Leu Gly Ile His Ile Asp Gly Asp Asp Met Lys Cys Thr Phe Gln  
 385 390 395 400  
 His Pro Ser Ser Ser Ser Gln Gln Ile Ala Tyr Phe Phe Asp Ser Cys  
 405 410 415  
 His Leu Leu Arg Leu Ile Arg Asn Ala Phe Gln Asn Phe Gln Ser Ile  
 420 425 430  
 Gln Phe Ile Asn Gly Ile Ala His Trp Gln His Leu Val Glu Leu Val  
 435 440 445  
 Ala Leu Glu Glu Gln Glu Leu Ser Asn Met Glu Arg Ile Pro Ser Thr  
 450 455 460  
 Leu Ala Asn Leu Lys Asn His Val Leu Lys Val Asn Ser Ala Thr Gln  
 465 470 475 480  
 Leu Phe Ser Glu Ser Val Ala Ser Ala Leu Glu Tyr Leu Leu Ser Leu  
 485 490 495  
 Asp Leu Pro Pro Phe Gln Asn Cys Ile Gly Thr Ile His Phe Leu Arg  
 500 505 510  
 Leu Ile Asn Asn Leu Phe Asp Ile Phe Asn Ser Arg Asn Cys Tyr Gly  
 515 520 525  
 Lys Gly Leu Lys Gly Pro Leu Leu Pro Glu Thr Tyr Ser Lys Ile Asn  
 530 535 540  
 His Val Leu Ile Glu Ala Lys Thr Ile Phe Val Thr Leu Ser Asp Thr  
 545 550 555 560  
 Ser Asn Asn Gln Ile Ile Lys Gly Lys Gln Lys Leu Gly Phe Leu Gly  
 565 570 575  
 Phe Leu Leu Asn Ala Glu Ser Leu Lys Trp Leu Tyr Gln Asn Tyr Val  
 580 585 590  
 Phe Pro Lys Val Met Pro Phe Pro Tyr Leu Leu Thr Tyr Lys Phe Ser  
 595 600 605  
 His Asp His Leu Glu Leu Phe Leu Lys Met Leu Arg Gln Val Leu Val  
 610 615 620  
 Thr Ser Ser Ser Pro Thr Cys Met Ala Phe Gln Lys Ala Tyr Tyr Asn  
 625 630 635 640  
 Leu Glu Thr Arg Tyr Lys Phe Gln Asp Glu Val Phe Leu Ser Lys Val

645 650 655  
 Ser Ile Phe Asp Ile Ser Ile Ala Arg Arg Lys Asp Leu Ala Leu Trp  
 660 665 670  
 Thr Val Gln Arg Gln Tyr Gly Val Ser Val Thr Lys Thr Val Phe His  
 675 680 685  
 Glu Glu Gly Ile Cys Gln Asp Trp Ser His Cys Ser Leu Ser Glu Ala  
 690 695 700  
 Leu Leu Asp Leu Ser Asp His Arg Arg Asn Leu Ile Cys Tyr Ala Gly  
 705 710 715 720  
 Tyr Val Ala Asn Lys Leu Ser Ala Leu Leu Thr Cys Glu Asp Cys Ile  
 725 730 735  
 Thr Ala Leu Tyr Ala Ser Asp Leu Lys Ala Ser Lys Ile Gly Ser Leu  
 740 745 750  
 Leu Phe Val Lys Lys Lys Asn Gly Leu His Phe Pro Ser Glu Ser Leu  
 755 760 765  
 Cys Arg Val Ile Asn Ile Cys Glu Arg Val Val Arg Thr His Ser Arg  
 770 775 780  
 Met Ala Ile Phe Glu Leu Val Ser Lys Gln Arg Glu Leu Tyr Leu Gln  
 785 790 795 800  
 Gln Lys Ile Leu Cys Glu Leu Ser Gly His Ile Asp Leu Phe Val Asp  
 805 810 815  
 Val Asn Lys His Leu Phe Asp Gly Glu Val Cys Ala Ile Asn His Phe  
 820 825 830  
 Val Lys Leu Leu Lys Asp Ile Ile Ile Cys Phe Leu Asn Ile Arg Ala  
 835 840 845  
 Lys Asn Val Ala Gln Asn Pro Leu Lys His His Ser Glu Arg Thr Asp  
 850 855 860  
 Met Lys Thr Leu Ser Arg Lys His Trp Ser Pro Val Gln Asp Tyr Lys  
 865 870 875 880  
 Cys Ser Ser Phe Ala Asn Thr Ser Ser Lys Phe Arg His Leu Leu Ser  
 885 890 895  
 Asn Asp Gly Tyr Pro Phe Lys  
 900

<210> 12  
 <211> 257  
 <212> PRT  
 <213> Homo sapiens

<400> 12  
 Met Pro Ala Arg Cys Val Ala Ala His Cys Gly Asn Thr Thr Lys Ser  
 1 5 10 15  
 Gly Lys Ser Leu Phe Arg Phe Pro Lys Asp Arg Ala Val Arg Leu Leu  
 20 25 30  
 Trp Asp Arg Phe Val Arg Gly Cys Arg Ala Asp Trp Tyr Gly Gly Asn  
 35 40 45  
 Asp Arg Ser Val Ile Cys Ser Asp His Phe Ala Pro Ala Cys Phe Asp  
 50 55 60  
 Val Ser Ser Val Ile Gln Lys Asn Leu Arg Phe Ser Gln Arg Leu Arg  
 65 70 75 80  
 Leu Val Ala Gly Ala Val Pro Thr Leu His Arg Val Pro Ala Pro Ala  
 85 90 95  
 Pro Lys Arg Gly Glu Glu Gly Asp Gln Ala Gly Arg Leu Asp Thr Arg  
 100 105 110  
 Gly Glu Leu Gln Ala Ala Arg His Ser Glu Ala Ala Pro Gly Pro Val  
 115 120 125  
 Ser Cys Thr Arg Pro Arg Ala Gly Lys Gln Ala Ala Ala Ser Gln Ile  
 130 135 140  
 Thr Cys Glu Asn Glu Leu Val Gln Thr Gln Pro His Ala Asp Asn Pro  
 145 150 155 160

Ser Asn Thr Val Thr Ser Val Pro Thr His Cys Glu Glu Gly Pro Val  
 165 170 175  
 His Lys Ser Thr Gln Ile Ser Leu Lys Arg Pro Arg His Arg Ser Val  
 180 185 190  
 Gly Ile Gln Ala Lys Val Lys Ala Phe Gly Lys Arg Leu Cys Asn Ala  
 195 200 205  
 Thr Thr Gln Thr Glu Glu Leu Trp Ser Arg Thr Ser Ser Leu Phe Asp  
 210 215 220  
 Ile Tyr Ser Ser Asp Ser Glu Thr Asp Thr Asp Trp Asp Ile Lys Ser  
 225 230 235 240  
 Glu Gln Ser Asp Leu Ser Tyr Met Ala Val Gln Val Lys Glu Glu Thr  
 245 250 255  
 Cys

<210> 13  
 <211> 314  
 <212> PRT  
 <213> Homo sapiens

<400> 13  
 Met Pro Gly Phe Thr Cys Cys Val Pro Gly Cys Tyr Asn Asn Ser His  
 1 5 10 15  
 Arg Asp Lys Ala Leu His Phe Tyr Thr Phe Pro Lys Asp Ala Glu Leu  
 20 25 30  
 Arg Arg Leu Trp Leu Lys Asn Val Ser Arg Ala Gly Val Ser Gly Cys  
 35 40 45  
 Phe Ser Thr Phe Gln Pro Thr Thr Gly His Arg Leu Cys Ser Val His  
 50 55 60  
 Phe Gln Gly Gly Arg Lys Thr Tyr Thr Val Arg Val Pro Thr Ile Phe  
 65 70 75 80  
 Pro Leu Arg Gly Val Asn Glu Arg Lys Val Ala Arg Arg Pro Ala Gly  
 85 90 95  
 Ala Ala Ala Ala Arg Arg Arg Gln Gln Gln Gln Gln Gln Gln Gln  
 100 105 110  
 Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln  
 115 120 125  
 Gln Gln Gln Gln Ser Ser Pro Ser Ala Ser Thr Ala Gln Thr Ala Gln  
 130 135 140  
 Leu Gln Pro Asn Leu Val Ser Ala Ser Ala Ala Val Leu Leu Thr Leu  
 145 150 155 160  
 Gln Ala Thr Val Asp Ser Ser Gln Ala Pro Gly Ser Val Gln Pro Ala  
 165 170 175  
 Pro Ile Thr Pro Thr Gly Glu Asp Val Lys Pro Ile Asp Leu Thr Val  
 180 185 190  
 Gln Val Glu Phe Ala Ala Ala Glu Gly Ala Ala Ala Ala Ala Ala Ala  
 195 200 205  
 Ser Glu Leu Gln Ala Ala Thr Ala Gly Leu Glu Ala Ala Glu Cys Pro  
 210 215 220  
 Met Gly Pro Gln Leu Val Val Gly Glu Glu Gly Phe Pro Asp Thr  
 225 230 235 240  
 Gly Ser Asp His Ser Tyr Ser Leu Ser Ser Gly Thr Thr Glu Glu Glu  
 245 250 255  
 Leu Leu Arg Lys Leu Asn Glu Gln Arg Asp Ile Leu Ala Leu Met Glu  
 260 265 270  
 Val Lys Met Lys Glu Met Lys Gly Ser Ile Arg His Leu Arg Leu Thr  
 275 280 285  
 Glu Ala Lys Leu Arg Glu Glu Leu Arg Glu Lys Asp Arg Leu Leu Ala  
 290 295 300  
 Met Ala Val Ile Arg Lys Lys His Gly Met



305

310

<210> 14  
 <211> 761  
 <212> PRT  
 <213> Homo sapiens

&lt;400&gt; 14

Met	Pro	Asn	Phe	Cys	Ala	Ala	Pro	Asn	Cys	Thr	Arg	Lys	Ser	Thr	Gln
1				5					10					15	
Ser	Asp	Leu	Ala	Phe	Phe	Arg	Phe	Pro	Arg	Asp	Pro	Ala	Arg	Cys	Gln
			20					25					30		
Lys	Trp	Val	Glu	Asn	Cys	Arg	Arg	Ala	Asp	Leu	Glu	Asp	Lys	Thr	Pro
		35					40					45			
Asp	Gln	Leu	Asn	Lys	His	Tyr	Arg	Leu	Cys	Ala	Lys	His	Phe	Glu	Thr
	50					55					60				
Ser	Met	Ile	Cys	Arg	Thr	Ser	Pro	Tyr	Arg	Thr	Val	Leu	Arg	Asp	Asn
65					70					75					80
Ala	Ile	Pro	Thr		Phe	Asp	Leu	Thr	Ser	His	Leu	Asn	Asn	Pro	His
				85					90					95	
Ser	Arg	His	Arg	Lys	Arg	Ile	Lys	Glu	Leu	Ser	Glu	Asp	Glu	Ile	Arg
			100					105					110		
Thr	Leu	Lys	Gln	Lys	Lys	Ile	Asp	Glu	Thr	Ser	Glu	Gln	Glu	Gln	Lys
		115					120					125			
His	Lys	Glu	Thr	Asn	Asn	Ser	Asn	Ala	Gln	Asn	Pro	Ser	Glu	Glu	Glu
	130					135					140				
Gly	Glu	Gly	Gln	Asp	Glu	Asp	Ile	Leu	Pro	Leu	Thr	Leu	Glu	Glu	Lys
145					150					155					160
Glu	Asn	Lys	Glu	Tyr	Leu	Lys	Ser	Leu	Phe	Glu	Ile	Leu	Ile	Leu	Met
				165					170					175	
Gly	Lys	Gln	Asn	Ile	Pro	Leu	Asp	Gly	His	Glu	Ala	Asp	Glu	Ile	Pro
			180					185					190		
Glu	Gly	Leu	Phe	Thr	Pro	Asp	Asn	Phe	Gln	Ala	Leu	Leu	Glu	Cys	Arg
	195						200					205			
Ile	Asn	Ser	Gly	Glu	Glu	Val	Leu	Arg	Lys	Arg	Phe	Glu	Thr	Thr	Ala
	210					215					220				
Val	Asn	Thr	Leu	Phe	Cys	Ser	Lys	Thr	Gln	Gln	Arg	Gln	Met	Leu	Glu
225					230					235					240
Ile	Cys	Glu	Ser	Cys	Ile	Arg	Glu	Glu	Thr	Leu	Arg	Glu	Val	Arg	Asp
				245					250					255	
Ser	His	Phe	Phe	Ser	Ile	Ile	Thr	Asp	Asp	Val	Val	Asp	Ile	Ala	Gly
			260					265					270		
Glu	Glu	His	Leu	Pro	Val	Leu	Val	Arg	Phe	Val	Asp	Glu	Ser	His	Asn
			275				280					285			
Leu	Arg	Glu	Glu	Phe	Ile	Gly	Phe	Leu	Pro	Tyr	Glu	Ala	Asp	Ala	Glu
	290					295					300				
Ile	Leu	Ala	Val	Lys	Phe	His	Thr	Met	Ile	Thr	Glu	Lys	Trp	Gly	Leu
305					310					315					320
Asn	Met	Glu	Tyr	Cys	Arg	Gly	Gln	Ala	Tyr	Ile	Val	Ser	Ser	Gly	Phe
				325					330					335	
Ser	Ser	Lys	Met	Lys	Val	Val	Ala	Ser	Arg	Leu	Leu	Glu	Lys	Tyr	Pro
			340					345					350		
Gln	Ala	Ile	Tyr	Thr	Leu	Cys	Ser	Ser	Cys	Ala	Leu	Asn	Met	Trp	Leu
		355					360					365			
Ala	Lys	Ser	Val	Pro	Val	Met	Gly	Val	Ser	Val	Ala	Leu	Gly	Thr	Ile
	370					375					380				
Glu	Glu	Val	Cys	Ser	Phe	Phe	His	Arg	Ser	Pro	Gln	Leu	Leu	Leu	Glu
385					390					395					400
Leu	Asp	Asn	Val	Ile	Ser	Val	Leu	Phe	Gln	Asn	Ser	Lys	Glu	Arg	Gly
				405					410					415	

Lys Glu Leu Lys Glu Ile Cys His Ser Gln Trp Thr Gly Arg His Asp  
                   420                  425                  430  
 Ala Phe Glu Ile Leu Val Glu Leu Gln Ala Leu Val Leu Cys Leu  
                   435                  440                  445  
 Asp Gly Ile Asn Ser Asp Thr Asn Ile Arg Trp Asn Asn Tyr Ile Ala  
                   450                  455                  460  
 Gly Arg Ala Phe Val Leu Cys Ser Ala Val Ser Asp Phe Asp Phe Ile  
                   465                  470                  475                  480  
 Val Thr Ile Val Val Leu Lys Asn Val Leu Ser Phe Thr Arg Ala Phe  
                   485                  490                  495  
 Gly Lys Asn Leu Gln Gly Gln Thr Ser Asp Val Phe Phe Ala Ala Gly  
                   500                  505                  510  
 Ser Leu Thr Ala Val Leu His Ser Leu Asn Glu Val Met Glu Asn Ile  
                   515                  520                  525  
 Glu Val Tyr His Glu Phe Trp Phe Glu Glu Ala Thr Asn Leu Ala Thr  
                   530                  535                  540  
 Lys Leu Asp Ile Gln Met Lys Leu Pro Gly Lys Phe Arg Arg Ala His  
                   545                  550                  555                  560  
 Gln Gly Asn Leu Glu Ser Gln Leu Thr Ser Glu Ser Tyr Tyr Lys Glu  
                   565                  570                  575  
 Thr Leu Ser Val Pro Thr Val Glu His Ile Ile Gln Glu Leu Lys Asp  
                   580                  585                  590  
 Ile Phe Ser Glu Gln His Leu Lys Ala Leu Lys Cys Leu Ser Leu Val  
                   595                  600                  605  
 Pro Ser Val Met Gly Gln Leu Lys Phe Asn Thr Ser Glu Glu His His  
                   610                  615                  620  
 Ala Asp Met Tyr Arg Ser Asp Leu Pro Asn Pro Asp Thr Leu Ser Ala  
                   625                  630                  635                  640  
 Glu Leu His Cys Trp Arg Ile Lys Trp Lys His Arg Gly Lys Asp Ile  
                   645                  650                  655  
 Glu Leu Pro Ser Thr Ile Tyr Glu Ala Leu His Leu Pro Asp Ile Lys  
                   660                  665                  670  
 Phe Phe Pro Asn Val Tyr Ala Leu Leu Lys Val Leu Cys Ile Leu Pro  
                   675                  680                  685  
 Val Met Lys Val Glu Asn Glu Arg Tyr Glu Asn Gly Arg Lys Arg Leu  
                   690                  695                  700  
 Lys Ala Tyr Leu Arg Asn Thr Leu Thr Asp Gln Arg Ser Ser Asn Leu  
                   705                  710                  715                  720  
 Ala Leu Leu Asn Ile Asn Phe Asp Ile Lys His Asp Leu Asp Leu Met  
                   725                  730                  735  
 Val Asp Thr Tyr Ile Lys Leu Tyr Thr Ser Lys Ser Glu Leu Pro Thr  
                   740                  745                  750  
 Asp Asn Ser Glu Thr Val Glu Asn Thr  
                   755                  760

&lt;210&gt; 15

&lt;211&gt; 38

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Consensus sequence for PAR4 binding domain of THAP

&lt;221&gt; UNSURE

&lt;222&gt; (1)...(38)

&lt;223&gt; Xaa = Any Amino Acid

&lt;400&gt; 15

Leu Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
 1                  5                  10                  15

Gln Arg Xaa Arg Arg Gln Xaa Arg Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
                   20                  25                  30  
 Xaa Xaa Xaa Gln Xaa Glu  
                   35

<210> 16  
 <211> 73  
 <212> PRT  
 <213> Sus scrofa

<400> 16  
 Met Val Gln Ser Cys Ser Ala Tyr Gly Cys Lys Asn Arg Tyr Asp Lys  
   1                  5                  10                  15  
 Asp Lys Pro Val Ser Phe His Lys Phe Pro Leu Thr Arg Pro Ser Leu  
                   20                  25                  30  
 Cys Lys Lys Trp Glu Ala Ala Val Arg Arg Lys Asn Phe Lys Pro Thr  
                   35                  40                  45  
 Lys Tyr Ser Ser Ile Cys Ser Glu His Phe Thr Pro Asp Cys Phe Lys  
                   50                  55                  60  
 Arg Glu Cys Asn Asn Lys Leu Leu Lys  
   65                  70

<210> 17  
 <211> 99  
 <212> PRT  
 <213> Sus scrofa

<400> 17  
 Met Val Lys Cys Cys Ser Ala Ile Gly Cys Ala Ser Arg Cys Leu Pro  
   1                  5                  10                  15  
 Asn Ser Lys Leu Lys Gly Leu Thr Phe His Val Phe Pro Thr Asp Glu  
                   20                  25                  30  
 Lys Val Lys Arg Lys Trp Val Leu Ala Met Lys Arg Leu Asp Val Asn  
                   35                  40                  45  
 Ala Ala Gly Met Trp Glu Pro Lys Lys Gly Asp Val Leu Cys Ser Arg  
                   50                  55                  60  
 His Phe Lys Lys Thr Asp Phe Asp Arg Thr Thr Pro Asn Ile Lys Leu  
   65                  70                  75                  80  
 Lys Pro Gly Val Ile Pro Ser Ile Phe Asp Ser Pro Ser His Leu Thr  
                   85                  90                  95  
 Gly Glu Glu

<210> 18  
 <211> 103  
 <212> PRT  
 <213> Sus scrofa

<400> 18  
 Met Pro Arg His Cys Ser Ala Ala Gly Cys Cys Thr Arg Asp Thr Arg  
   1                  5                  10                  15  
 Glu Thr Arg Asn Arg Gly Ile Ser Phe His Arg Leu Pro Lys Asp  
                   20                  25                  30  
 Asn Pro Arg Arg Gly Leu Trp Leu Ala Asn Cys Gln Arg Leu Asp Pro  
                   35                  40                  45  
 Ser Gly Gln Gly Leu Trp Asp Pro Ala Ser Glu Tyr Ile Tyr Phe Cys  
                   50                  55                  60  
 Ser Lys His Phe Glu Glu Asn Cys Phe Glu Leu Val Gly Ile Ser Gly

```
<210> 19
<211> 99
<212> PRT
<213> Sus scrofa
```

```
<210> 20
<211> 92
<212> PRT
<213> Bos taurus
```

```
<210> 21
<211> 75
<212> PRT
<213> Bos taurus
```

-15-

Leu Val Gly Ile Ser Gly Tyr His Arg Leu Lys Glu Gly Ala Val Pro  
 50 55 60  
 Thr Ile Phe Glu Ser Phe Ser Lys Leu Arg Arg  
 65 70 75

<210> 22  
 <211> 91  
 <212> PRT  
 <213> Mus musculus

<400> 22  
 Met Val Gln Ser Cys Ser Ala Tyr Gly Cys Lys Asn Arg Tyr Asp Lys  
 1 5 10 15  
 Asp Lys Pro Val Ser Phe His Lys Phe Pro Leu Thr Arg Pro Ser Leu  
 20 25 30  
 Cys Lys Gln Trp Glu Ala Ala Val Lys Arg Lys Asn Phe Lys Pro Thr  
 35 40 45  
 Lys Tyr Ser Ser Ile Cys Ser Glu His Phe Thr Pro Asp Cys Phe Lys  
 50 55 60  
 Arg Glu Cys Asn Asn Lys Leu Leu Lys Glu Asn Ala Val Pro Thr Ile  
 65 70 75 80  
 Phe Leu Tyr Ile Glu Pro His Glu Lys Lys Glu  
 85 90

<210> 23  
 <211> 90  
 <212> PRT  
 <213> Mus musculus

<400> 23  
 Met Pro Thr Asn Cys Ala Ala Ala Gly Cys Ala Ala Thr Tyr Asn Lys  
 1 5 10 15  
 His Ile Asn Ile Ser Phe His Arg Phe Pro Leu Asp Pro Lys Arg Arg  
 20 25 30  
 Lys Glu Trp Val Arg Leu Val Arg Arg Lys Asn Phe Val Pro Gly Lys  
 35 40 45  
 His Thr Phe Leu Cys Ser Lys His Phe Glu Ala Ser Cys Phe Asp Leu  
 50 55 60  
 Thr Gly Gln Thr Arg Arg Leu Lys Met Asp Ala Val Pro Thr Ile Phe  
 65 70 75 80  
 Asp Phe Cys Thr His Ile Lys Ser Leu Lys  
 85 90

<210> 24  
 <211> 92  
 <212> PRT  
 <213> Mus musculus

<400> 24  
 Met Pro Lys Ser Cys Ala Ala Arg Gln Cys Cys Asn Arg Tyr Ser Ser  
 1 5 10 15  
 Arg Arg Lys Gln Leu Thr Phe His Arg Phe Pro Phe Ser Arg Pro Glu  
 20 25 30  
 Leu Leu Arg Glu Trp Val Leu Asn Ile Gly Arg Ala Asp Phe Lys Pro  
 35 40 45  
 Lys Gln His Thr Val Ile Cys Ser Glu His Phe Arg Pro Glu Cys Phe  
 50 55 60  
 Ser Ala Phe Gly Asn Arg Lys Asn Leu Lys His Asn Ala Val Pro Thr

```
<210> 25
<211> 95
<212> PRT
<213> Mus musculus
```

```
<210> 26
<211> 52
<212> PRT
<213> Mus musculus
```

```
<210> 27
<211> 103
<212> PRT
<213> Mus musculus
```

-17-

<210> 28  
 <211> 90  
 <212> PRT  
 <213> Mus musculus

<400> 28  
 Met Pro Gly Phe Thr Cys Cys Val Pro Gly Cys Tyr Asn Asn Ser His  
 1 5 10 15  
 Arg Asp Lys Ala Leu His Phe Tyr Thr Phe Pro Lys Asp Ala Glu Leu  
 20 25 30  
 Arg Arg Leu Trp Leu Lys Asn Val Ser Arg Ala Gly Val Ser Gly Cys  
 35 40 45  
 Phe Ser Thr Phe Gln Pro Thr Thr Gly His Arg Leu Cys Ser Val His  
 50 55 60  
 Phe Gln Gly Gly Arg Lys Thr Tyr Thr Val Arg Val Pro Thr Ile Phe  
 65 70 75 80  
 Pro Leu Arg Gly Val Asn Glu Arg Lys Val  
 85 90

<210> 29  
 <211> 96  
 <212> PRT  
 <213> Mus musculus

<400> 29  
 Met Pro Asn Phe Cys Ala Ala Pro Asn Cys Thr Arg Lys Ser Thr Gln  
 1 5 10 15  
 Ser Asp Leu Ala Phe Phe Arg Phe Pro Arg Asp Pro Ala Arg Cys Gln  
 20 25 30  
 Lys Trp Val Glu Asn Cys Arg Arg Ala Asp Leu Glu Asp Lys Thr Pro  
 35 40 45  
 Asp Gln Leu Asn Lys His Tyr Arg Leu Cys Ala Lys His Phe Glu Thr  
 50 55 60  
 Ser Met Ile Cys Arg Thr Ser Pro Tyr Arg Thr Val Leu Arg Asp Asn  
 65 70 75 80  
 Ala Ile Pro Thr Ile Phe Asp Leu Thr Ser His Leu Asn Asn Pro His  
 85 90 95

<210> 30  
 <211> 24  
 <212> PRT  
 <213> Rattus norvegicus

<400> 30  
 Met Pro Thr Asn Cys Ala Ala Ala Gly Cys Ala Ala Thr Tyr Asn Lys  
 1 5 10 15  
 His Ile Asn Ile Ser Phe His Arg  
 20

<210> 31  
 <211> 85  
 <212> PRT  
 <213> Rattus norvegicus

<400> 31  
 Arg Gln Cys Cys Asn Arg Tyr Ser Ser Arg Arg Lys Gln Leu Thr Phe

```

1           5           10           15
His Arg Phe Pro Phe Ser Arg Pro Glu Leu Leu Arg Glu Trp Val Leu
20           25           30
Asn Ile Gly Arg Ala Asp Phe Lys Pro Lys Gln His Thr Val Ile Cys
35           40           45
Ser Glu His Phe Arg Pro Glu Cys Phe Ser Ala Phe Gly Asn Arg Lys
50           55           60
Asn Leu Lys His Asn Ala Val Pro Thr Val Phe Ala Phe Gln Asn Pro
65           70           75           80
Ala Gln Val Cys Pro
85

```

<210> 32  
 <211> 70  
 <212> PRT  
 <213> Rattus norvegicus

```

<400> 32
Arg Phe Pro Leu Lys Asp Ser Lys Arg Leu Ile Gln Trp Leu Lys Ala
1           5           10           15
Val Gln Arg Asp Asn Trp Thr Pro Thr Lys Tyr Ser Phe Leu Cys Ser
20           25           30
Glu His Phe Thr Lys Asp Ser Phe Ser Lys Arg Leu Glu Asp Gln His
35           40           45
Arg Leu Leu Lys Pro Thr Ala Val Pro Ser Ile Phe His Leu Ser Glu
50           55           60
Lys Lys Arg Gly Ala Gly
65           70

```

<210> 33  
 <211> 55  
 <212> PRT  
 <213> Rattus norvegicus

```

<400> 33
Met Val Lys Cys Cys Ser Ala Ile Gly Cys Ala Ser Arg Cys Leu Pro
1           5           10           15
Asn Ser Lys Leu Lys Gly Leu Thr Phe His Val Phe Pro Thr Asp Glu
20           25           30
Asn Ile Lys Arg Lys Trp Val Leu Ala Met Lys Arg Leu Asp Val Asn
35           40           45
Thr Ala Gly Ile Trp Glu Pro
50           55

```

<210> 34  
 <211> 103  
 <212> PRT  
 <213> Rattus norvegicus

```

<400> 34
Met Pro Arg His Cys Ser Ala Ala Gly Cys Cys Thr Arg Asp Thr Arg
1           5           10           15
Glu Thr Arg Asn Arg Gly Ile Ser Phe His Arg Leu Pro Lys Lys Asp
20           25           30
Asn Pro Arg Arg Gly Leu Trp Leu Ala Asn Cys Gln Arg Leu Asp Pro
35           40           45
Ser Gly Gln Gly Leu Trp Asp Pro Thr Ser Glu Tyr Ile Tyr Phe Cys
50           55           60

```



Ser Lys His Phe Glu Glu Asn Cys Phe Glu Leu Val Gly Ile Ser Gly  
 65 70 75 80  
 Tyr His Arg Leu Lys Glu Gly Ala Val Pro Thr Ile Phe Glu Ser Phe  
 85 90 95  
 Ser Lys Leu Arg Arg Thr Ala  
 100

<210> 35  
 <211> 90  
 <212> PRT  
 <213> Rattus norvegicus

<400> 35  
 Met Pro Gly Phe Thr Cys Cys Val Pro Gly Cys Tyr Asn Asn Ser His  
 1 5 10 15  
 Arg Asp Lys Ala Leu His Phe Tyr Thr Phe Pro Lys Asp Ala Glu Leu  
 20 25 30  
 Arg Arg Leu Trp Leu Lys Asn Val Ser Arg Ala Gly Val Ser Gly Cys  
 35 40 45  
 Phe Ser Thr Phe Gln Pro Thr Thr Gly His Arg Leu Cys Ser Val His  
 50 55 60  
 Phe Gln Gly Gly Arg Lys Thr Tyr Thr Val Arg Val Pro Thr Ile Phe  
 65 70 75 80  
 Pro Leu Arg Gly Val Asn Glu Arg Lys Val  
 85 90

<210> 36  
 <211> 96  
 <212> PRT  
 <213> Rattus norvegicus

<400> 36  
 Met Pro Asn Phe Cys Ala Ala Pro Asn Cys Thr Arg Lys Ser Thr Gln  
 1 5 10 15  
 Ser Asp Leu Ala Phe Phe Arg Phe Pro Arg Asp Pro Ala Arg Cys Gln  
 20 25 30  
 Lys Trp Val Glu Asn Cys Arg Arg Ala Asp Leu Glu Asp Lys Thr Pro  
 35 40 45  
 Asp Gln Leu Asn Lys His Tyr Arg Leu Cys Ala Lys His Phe Glu Thr  
 50 55 60  
 Ser Met Ile Cys Arg Thr Ser Pro Tyr Arg Thr Val Leu Arg Asp Asn  
 65 70 75 80  
 Ala Ile Pro Thr Ile Phe Asp Leu Thr Ser His Leu Asn Asn Pro His  
 85 90 95

<210> 37  
 <211> 94  
 <212> PRT  
 <213> Gallus gallus

<400> 37  
 Met Val Ile Cys Cys Ala Ala Ala Asn Cys Ser Asn Arg Gln Gly Lys  
 1 5 10 15  
 Ala Leu Arg Gly Ala Val Ser Phe His Arg Phe Pro Leu Lys Asp Ser  
 20 25 30  
 Lys Arg Leu Ile Gln Trp Leu Lys Ala Val Gln Arg Asp Asn Trp Thr  
 35 40 45  
 Pro Thr Lys Tyr Ser Phe Leu Cys Ser Glu His Phe Thr Lys Asp Ser

50 55 60  
 Phe Ser Arg Arg Leu Glu Asp Gln His Arg Leu Leu Lys Pro Thr Ala  
 65 70 75 80  
 Val Pro Thr Ile Phe Gln Leu Ala Glu Lys Lys Arg Asp Asn  
 85 90

<210> 38  
 <211> 94  
 <212> PRT  
 <213> Gallus gallus

<400> 38  
 Met Pro Arg Tyr Cys Ala Ala Ser Tyr Cys Lys Asn Arg Gly Gly Gln  
 1 5 10 15  
 Ser Ala Arg Asp Gln Arg Lys Leu Ser Phe Tyr Pro Phe Pro Leu His  
 20 25 30  
 Asp Lys Glu Arg Leu Glu Lys Trp Leu Arg Asn Met Lys Arg Asp Ala  
 35 40 45  
 Trp Thr Pro Ser Lys His Gln Leu Leu Cys Ser Asp His Phe Thr Pro  
 50 55 60  
 Asp Ser Leu Asp Val Arg Trp Gly Ile Arg Tyr Leu Lys His Thr Ala  
 65 70 75 80  
 Val Pro Thr Ile Phe Ser Ser Pro Asp Asp Glu Glu Lys Gly  
 85 90

<210> 39  
 <211> 102  
 <212> PRT  
 <213> Gallus gallus

<400> 39  
 Met Pro Arg His Cys Ser Ala Ala Gly Cys Cys Thr Arg Asp Thr Arg  
 1 5 10 15  
 Glu Thr Arg Ser Arg Gly Ile Ser Phe His Arg Leu Pro Lys Lys Asp  
 20 25 30  
 Asn Pro Arg Arg Ala Leu Trp Leu Glu Asn Ser Arg Arg Asp Ala  
 35 40 45  
 Ser Gly Glu Gly Arg Trp Asp Pro Ala Ser Lys Tyr Ile Tyr Phe Cys  
 50 55 60  
 Ser Gln His Phe Glu Lys Ser Cys Phe Glu Ile Val Gly Phe Ser Gly  
 65 70 75 80  
 Tyr His Arg Leu Lys Glu Gly Ala Val Pro Thr Val Phe Glu Ser Thr  
 85 90 95  
 Ser Pro Arg Pro Pro Arg  
 100

<210> 40  
 <211> 27  
 <212> PRT  
 <213> Gallus gallus

<400> 40  
 Met Thr Arg Ser Cys Ser Ala Leu Gly Cys Ser Ala Arg Asp Asn Gly  
 1 5 10 15  
 Arg Ser Arg Glu Arg Gly Ile Ser Phe His Gln  
 20 25

<210> 41  
 <211> 90  
 <212> PRT  
 <213> Xenopus laevi

<400> 41  
 Met Val Gln Ser Cys Ser Ala Tyr Gly Cys Lys Asn Arg Tyr Asp Lys  
 1 5 10 15  
 Asp Arg Pro Ile Ser Phe His Lys Phe Pro Leu Lys Arg Pro Leu Leu  
 20 25 30  
 Cys Lys Lys Trp Glu Ala Ala Val Arg Arg Ala Asp Phe Lys Pro Thr  
 35 40 45  
 Lys Tyr Ser Ser Ile Cys Ser Asp His Phe Thr Ala Asp Cys Phe Lys  
 50 55 60  
 Arg Glu Cys Asn Asn Lys Leu Leu Lys Asp Asn Ala Val Pro Thr Val  
 65 70 75 80  
 Phe Ala Leu Ala Glu Ile Lys Lys Lys Met  
 85 90

<210> 42  
 <211> 103  
 <212> PRT  
 <213> Xenopus laevi

<400> 42  
 Met Pro Arg His Cys Ser Ala Leu Gly Cys Thr Thr Arg Asp Ser Arg  
 1 5 10 15  
 Gln Thr Arg Asn Asn Ile Ser Phe His Arg Leu Pro Arg Lys Asp  
 20 25 30  
 Asp Pro Arg Arg Asn Leu Trp Ile Ala Asn Cys Gln Arg Thr Asp Pro  
 35 40 45  
 Ser Gly Lys Gly Leu Trp Asp Pro Ser Ser Asp Tyr Val Tyr Phe Cys  
 50 55 60  
 Ser Lys His Phe Glu Lys Ser Cys Phe Glu Val Val Gly Thr Ser Gly  
 65 70 75 80  
 Tyr His Arg Leu Lys Glu Asp Ala Val Pro Thr Leu Phe Leu Ser Ser  
 85 90 95  
 Ala Lys Leu Arg Arg Ala Ala  
 100

<210> 43  
 <211> 90  
 <212> PRT  
 <213> Xenopus laevi

<400> 43  
 Met Val Arg Ser Cys Ser Ala Ala Asn Cys Val Asn Arg Gln Thr Ala  
 1 5 10 15  
 Leu Asn Lys Arg Lys Gly Ile Thr Phe His Arg Phe Pro Lys Glu Gln  
 20 25 30  
 Ala Arg Arg Gln Leu Trp Ile Thr Ala Val Thr His Ser His Ala Ala  
 35 40 45  
 Val Gly Thr Asp Trp Thr Pro Ser Ile His Ser Ser Leu Cys Ser Gln  
 50 55 60  
 His Phe Asn Asn Thr Gln Phe Asp Arg Thr Gly Gln Thr Val Arg Leu  
 65 70 75 80  
 Arg Asp Ser Ala Val Pro Thr Val Phe Ser  
 85 90

<210> 44  
 <211> 99  
 <212> PRT  
 <213> *Xenopus laevis*

<400> 44  
 Met Pro Val Ser Cys Ala Ala Ser Gly Cys Lys Ser Arg Tyr Thr Met  
 1 5 10 15  
 Asp Ala Arg Glu Lys Gly Ile Thr Phe His Arg Phe Pro Arg Ser Asn  
 20 25 30  
 Pro Thr Leu Leu Glu Lys Trp Arg Leu Ala Met Arg Arg Ser Thr Arg  
 35 40 45  
 Asn Gly Glu Leu Trp Met Pro Ser Arg Tyr Gln Arg Leu Cys Ser Leu  
 50 55 60  
 His Phe Lys Gln Cys Cys Phe Asp Thr Thr Gly Gln Thr Lys Arg Leu  
 65 70 75 80  
 Arg Glu Asp Val Ile Pro Thr Ile Phe Asp Phe Pro Glu Glu Thr His  
 85 90 95  
 Val Ile Phe

<210> 45  
 <211> 90  
 <212> PRT  
 <213> *Xenopus laevis*

<400> 45  
 Met Pro Ala Cys Ala Ala Ile Asn Cys Thr Ser Arg Gln Thr Arg Gly  
 1 5 10 15  
 Cys Gly Lys Ser Phe His Lys Phe Pro His Gly Arg Pro Glu Val Leu  
 20 25 30  
 Lys Lys Trp Val Met Asn Met Arg Arg Asp Lys Phe Lys Pro Ser Ser  
 35 40 45  
 Lys Ala Val Leu Cys Ser Asp His Phe Glu Glu Phe Cys Phe Asp Arg  
 50 55 60  
 Thr Gly Gln Thr Ile Arg Leu Arg Thr Asp Ala Val Pro Thr Val Phe  
 65 70 75 80  
 Thr Phe Pro Gly Lys Met Lys Lys Asp Arg  
 85 90

<210> 46  
 <211> 105  
 <212> PRT  
 <213> *Xenopus laevis*

<400> 46  
 Met Pro His Cys Val Val Ser Asn Cys Val His Phe Asn Tyr Lys Lys  
 1 5 10 15  
 Ser Asn Leu His Gly Val Ala Leu His Pro Phe Pro Asn Asp Leu Ser  
 20 25 30  
 Arg Ile Lys Leu Trp Leu Gln Gln Ile Gly Leu Thr Thr Asp Glu Ile  
 35 40 45  
 Asp Tyr Leu Ala Gln Lys Val Val Glu Gly Lys Arg Lys Lys Thr Asp  
 50 55 60  
 Ser His Arg Met Cys Ser Ala His Phe Thr Pro Asn Cys Tyr Ile Val  
 65 70 75 80  
 Gln Asp Ala Lys Leu Val Leu Arg Ser Asp Ala Ile Pro Thr Met Phe  
 85 90 95

Pro Gly Leu Ser Ser Ser Thr Thr Asn  
100 105

<210> 47  
<211> 104  
<212> PRT  
<213> *Xenopus laevis*

<400> 47  
Met Pro Lys Cys Ile Val Thr Lys Cys Pro His Lys Thr Gly Gln Lys  
1 5 10 15  
Glu Leu Tyr Pro Ser Val Ile Leu His Pro Phe Pro Gly Asn Ile Glu  
20 25 30  
Lys Ile Lys Gln Trp Leu Leu Gln Thr Gly Glu Asp Tyr Gly Asp Tyr  
35 40 45  
Glu Val Phe Ala Glu Lys Val Leu Glu Ala Lys Lys Thr Asp Ala Tyr  
50 55 60  
Arg Ile Cys Ser Arg His Phe Ala Glu Asp Gln Tyr Val Lys Arg Gly  
65 70 75 80  
Pro Arg Lys Leu Leu Ser Lys Asp Ala Val Pro Thr Ile Phe Ser Asn  
85 90 95  
Leu His Pro Leu Ile Gln Leu His  
100

<210> 48  
<211> 102  
<212> PRT  
<213> *Xenopus laevis*

<400> 48  
Met Pro Arg Cys Val Val Lys Asn Cys Pro His Trp Thr Gly Lys Lys  
1 5 10 15  
Gly Ser Gln Val Ile Leu His Gly Phe Pro Asn Asn Ser Arg Leu Ile  
20 25 30  
Lys Leu Trp Leu Ser Gln Thr Lys Gln Asp Phe Gly Asp Val Glu Asp  
35 40 45  
Phe Thr Gln Lys Ile Leu Glu Gly Lys Lys Asn Asp Leu Tyr Arg Leu  
50 55 60  
Cys Ser Lys His Phe Thr Asn Asp Ser Tyr Glu Ile Arg Gly Thr Lys  
65 70 75 80  
Arg Phe Leu Lys Tyr Gly Ala Val Pro Thr Val Phe Glu Asp Thr Pro  
85 90 95  
Pro Leu Lys Arg Arg Lys  
100

<210> 49  
<211> 104  
<212> PRT  
<213> *Xenopus laevis*

<400> 49  
Met Pro Asn Cys Ile Val Lys Asp Cys Arg His Lys Ser Gly Gln Lys  
1 5 10 15  
Ile Gln Asn Pro Asp Val Val Leu His Pro Phe Pro Asn Asn Ile Asn  
20 25 30  
Met Ile Lys Asn Trp Leu Leu Gln Thr Gly Gln Asp Phe Gly Asp Ile  
35 40 45  
Asp Val Leu Ala Asp Lys Ile Leu Lys Gly Lys Lys Thr Ala Asn Phe

50                                      55                                      60  
 Arg Met Cys Ser Cys His Phe Thr Arg Asp Ser Tyr Met Ala Arg Gly  
 65                                      70                                      75                                      80  
 Ser Lys Thr Thr Leu Lys Pro Asn Ala Ile Pro Thr Ile Phe Pro Val  
                                     85                                      90                                      95  
 Ile Leu Pro Thr Thr Val Pro Ser  
                                     100

<210> 50  
 <211> 99  
 <212> PRT  
 <213> *Xenopus laevi*

<400> 50  
 Met Pro Lys Cys Phe Val Gln Ser Cys Pro His Tyr Thr Gly Arg Asn  
 1                                      5                                      10                                      15  
 Gly Lys Pro Asp Asn Val Ile Leu His Thr Phe Pro Arg Cys Lys Lys  
                                     20                                      25                                      30  
 Gln Val Gln Val Trp Leu Ser Arg Thr Gly Glu Arg Tyr Glu Asn Met  
                                     35                                      40                                      45  
 Ala Glu Phe Val Thr Tyr Ile Thr Gln Arg Cys Ser Asn Phe Arg Met  
                                     50                                      55                                      60  
 Cys Ser Glu His Phe Thr Asp Asp Cys Tyr Ile Thr Val Glu Gly Lys  
 65                                      70                                      75                                      80  
 Arg Arg Leu Met Glu Asn Ser Ala Pro Thr Ile Phe Lys Thr Thr Phe  
                                     85                                      90                                      95  
 Arg Gln Asn

<210> 51  
 <211> 104  
 <212> PRT  
 <213> *Xenopus laevi*

<400> 51  
 Met Thr Lys Cys Ile Val Lys Gly Cys Arg His Thr Thr Gly Gln Lys  
 1                                      5                                      10                                      15  
 Leu Lys Phe Pro His Ile Val Met His Ala Phe Pro Ser Asn Leu Lys  
                                     20                                      25                                      30  
 Met Ile Lys Val Trp Leu Lys Gln Thr Gly Gln Tyr Gly Asn Asn Leu  
                                     35                                      40                                      45  
 Glu Glu Met Ala Leu Lys Val Leu Gly Gly Lys Lys Ser Asp Ser Tyr  
                                     50                                      55                                      60  
 Arg Leu Cys Ser Ala His Phe Thr Val Asp Ser Tyr Ala Leu Arg Arg  
 65                                      70                                      75                                      80  
 Ser Lys Asn Met Leu Lys Lys Asp Ala Phe Pro Thr Leu Phe Gly Gln  
                                     85                                      90                                      95  
 Asn Gln Ile Asn Ala Ala Asn Val  
                                     100

<210> 52  
 <211> 84  
 <212> PRT  
 <213> *Xenopus laevi*

<400> 52  
 Met Pro Lys Cys Ile Val Ile His Cys Pro His Ser Cys Ser Lys Lys  
 1                                      5                                      10                                      15

Val Thr Lys Asn Thr Gly Val Val Met His Thr Phe Pro Phe Asn Leu  
                   20                  25                  30  
 Asp Arg Ile Lys Asn Trp Leu Leu Ser Ile Asp Gln Asn Phe Gly Asn  
                   35                  40                  45  
 Ile Asp Thr Leu Ala Asn Arg Ile Leu Glu Glu Lys Lys Lys His Ser  
                   50                  55                  60  
 Asp Leu Tyr Arg Leu Cys Ser Glu His Phe Thr Pro Gln Cys Tyr Ile  
                   65                  70                  75                  80  
 Ser Thr Gly Glu

<210> 53  
 <211> 104  
 <212> PRT  
 <213> Xenopus laevi

<400> 53  
 Met Pro Ser Cys Ile Val Lys Gly Cys Pro His Arg Thr Gly Gln Lys  
   1                  5                  10                  15  
 Asp Lys Phe Pro Asn Val Thr Leu His Asn Phe Pro Lys Thr Ile Pro  
                   20                  25                  30  
 Lys Ile Lys Asn Trp Leu Trp Gln Thr Gly Gln Tyr Gly Glu Asp Ser  
                   35                  40                  45  
 Asp Ala Ile Ala Glu Glu Ile Leu Gln Gly Leu Lys Thr Cys Arg His  
                   50                  55                  60  
 Arg Met Cys Ser Met His Phe Ser Glu Asn Cys Phe Ile Thr Leu Gly  
                   65                  70                  75                  80  
 Ser Lys Arg Val Leu Thr Arg Asn Ala Val Pro Thr Ile Phe Lys Pro  
                   85                  90                  95  
 Gln Thr Thr Pro Ala Ile Leu Ala  
                   100

<210> 54  
 <211> 104  
 <212> PRT  
 <213> Xenopus laevi

<400> 54  
 Met Pro Lys Cys Ile Leu Asn Gly Cys Pro Tyr Arg Thr Gly Gln Lys  
   1                  5                  10                  15  
 Leu Lys Phe Pro Asp Ile Val Leu His Pro Phe Pro Lys Ser Met Glu  
                   20                  25                  30  
 Met Ile Arg Asn Trp Leu Phe Gln Thr Gly Gln His Ala Glu Asp Val  
                   35                  40                  45  
 Glu Ser Leu Ser Gln Arg Ile Tyr Gln Gly Leu Lys Thr Ser Asn Phe  
                   50                  55                  60  
 Arg Met Cys Ser Lys His Phe Thr Gln Asp Cys Tyr Met Gln Val Gly  
                   65                  70                  75                  80  
 Ser Arg Lys Cys Leu Lys Pro Asn Ala Val Pro Thr Val Phe Glu Ser  
                   85                  90                  95  
 Tyr Asn Val Pro Val Thr Thr Phe  
                   100

<210> 55  
 <211> 105  
 <212> PRT  
 <213> Xenopus laevi

&lt;400&gt; 55

```

Asn Asn Ala Ser Cys Ile Val Arg Gly Cys His His Ser Thr Ala Arg
 1           5           10           15
Lys Cys Leu Ser Pro Gly Ile Ala Leu His Gly Phe Pro Asn Asn Leu
          20           25           30
Ser Arg Ile Lys Gln Trp Leu Val Asn Ile Gly Gln Asn Val Gly Asp
          35           40           45
Ile Asp Asp Phe Ala Gln Lys Val Leu Asp Gly Lys Lys Gln Asn Ser
          50           55           60
Tyr Arg Ile Cys Ser Ala His Phe Ser Ser Asp Cys Phe Val Gln Phe
          65           70           75           80
Gly Tyr Ser Lys Gly Leu Lys Ala Asp Ala Val Pro Thr Ile Phe Ala
          85           90           95
Trp Asn Thr Pro Glu Ser Arg Gly Arg
          100           105

```

&lt;210&gt; 56

&lt;211&gt; 107

&lt;212&gt; PRT

<213> *Xenopus laevi*

&lt;400&gt; 56

```

Met Pro Ser Cys Ile Val Lys Gly Cys Arg His Lys Ser Gly Gln Lys
 1           5           10           15
Val Leu Tyr Pro Asp Val Val Leu His Ser Phe Pro Asn Asn Ile His
          20           25           30
Met Ile Lys Asn Trp Leu Leu Gln Thr Gly Gln Val Phe Gly Asp Ile
          35           40           45
Asp Ala Phe Ala Glu Lys Val Leu Lys Gly Asn Lys Thr Ser Ala Phe
          50           55           60
Arg Met Cys Ser Arg His Phe Thr Arg Asp Ser Tyr Met Ala Lys Gly
          65           70           75           80
Ser Lys Ile Thr Leu Lys Pro Asn Ala Val Pro Thr Ile Phe Asn Thr
          85           90           95
Leu Pro Pro Ala Ala Val Pro Ser Leu Met
          100           105

```

&lt;210&gt; 57

&lt;211&gt; 91

&lt;212&gt; PRT

<213> *Danio rerio*

&lt;400&gt; 57

```

Met Val Gln Ser Cys Ser Ala Tyr Gly Cys Asn Asn Arg Tyr Gln Lys
 1           5           10           15
Asp Arg Ile Ile Ser Phe His Lys Phe Pro Leu Ala Arg Pro Glu Val
          20           25           30
Cys Val Gln Trp Val Ser Ala Met Ser Arg Arg Asn Phe Lys Pro Thr
          35           40           45
Lys Tyr Ser Asn Ile Cys Ser Gln His Phe Thr Ser Asp Cys Phe Lys
          50           55           60
Gln Glu Cys Asn Asn Arg Val Leu Lys Asp Asn Ala Val Pro Ser Leu
          65           70           75           80
Phe Thr Leu Gln Thr Gln Asp Pro Phe Ser Ala
          85           90

```

&lt;210&gt; 58

&lt;211&gt; 103



&lt;212&gt; PRT

&lt;213&gt; Danio rerio

&lt;400&gt; 58

```

Met Pro Arg His Cys Ser Ala Val Gly Cys Lys Ser Arg Asp Thr Lys
 1           5           10           15
Asp Val Arg Lys Ser Gly Ile Thr Phe His Arg Leu Pro Lys Gly
      20           25           30
Asn Pro Arg Arg Thr Thr Trp Ile Asn Ser Arg Arg Lys Gly Pro
      35           40           45
Glu Gly Lys Gly Gln Trp Asp Pro Gln Ser Gly Phe Ile Tyr Phe Cys
      50           55           60
Ser Lys His Phe Thr Pro Asp Ser Phe Glu Leu Ser Gly Val Ser Gly
      65           70           75           80
Tyr His Arg Leu Lys Asp Asp Ala Ile Pro Thr Val Phe Glu Ile Glu
      85           90           95
Pro His Lys Lys Gly Thr Ala
      100

```

&lt;210&gt; 59

&lt;211&gt; 90

&lt;212&gt; PRT

&lt;213&gt; Danio rerio

&lt;400&gt; 59

```

Met Pro Gly Phe Thr Cys Cys Val Pro Gly Cys Tyr Asn Asn Ser His
 1           5           10           15
Arg Asp Arg Asp Leu Arg Phe Tyr Thr Phe Pro Lys Asp Pro Thr Gln
      20           25           30
Arg Glu Ile Trp Leu Lys Asn Ile Ser Arg Ala Gly Val Ser Gly Cys
      35           40           45
Phe Ser Thr Phe Gln Pro Thr Thr Gly His Arg Val Cys Ser Val His
      50           55           60
Phe Pro Gly Gly Arg Lys Thr Tyr Thr Ile Arg Val Pro Thr Leu Phe
      65           70           75           80
Pro Leu Arg Gly Val Asn Glu Arg Arg Ser
      85           90

```

&lt;210&gt; 60

&lt;211&gt; 96

&lt;212&gt; PRT

&lt;213&gt; Danio rerio

&lt;400&gt; 60

```

Met Pro Asn Phe Cys Ala Ala Leu Asn Cys Ser Arg Asn Ser Thr His
 1           5           10           15
Ser Val Leu Ala Phe Phe Arg Phe Pro Arg Asp Pro Glu Arg Cys Lys
      20           25           30
Lys Trp Val Glu Asn Cys Ser Arg Ser Asp Leu Lys Asp Lys Thr Pro
      35           40           45
Asp His Leu Asn Lys Tyr His Arg Leu Cys Ala Arg His Phe Glu Pro
      50           55           60
Asn Leu Ile Thr Lys Thr Ser Pro Phe Arg Thr Val Leu Lys Asp Ser
      65           70           75           80
Ala Val Pro Thr Ile Phe Asp Asn Pro Phe Lys Arg Ser Asn Asn Glu
      85           90           95

```

&lt;210&gt; 61

<211> 99  
 <212> PRT  
 <213> Danio rerio

<400> 61  
 Met Pro Tyr Lys Cys Val Ala Tyr Gly Cys Gly Lys Ile Ser Gly Gln  
 1 5 10 15  
 Asn Val Ser Met Phe Arg Phe Pro Lys Asp Pro Glu Glu Phe Ser Lys  
 20 25 30  
 Trp Gln Arg Gln Val Gln Lys Thr Arg Arg Asn Trp Leu Ala Asn Thr  
 35 40 45  
 Tyr Ser His Leu Cys Asn Glu His Phe Thr Lys Asp Cys Phe Glu Pro  
 50 55 60  
 Lys Thr Tyr Val Thr Ala Lys Ala Ser Gly Phe Lys Arg Leu Lys Leu  
 65 70 75 80  
 Lys Asp Gly Ala Val Pro Thr Val Phe Ile Arg Arg Arg Cys Arg Lys  
 85 90 95  
 Cys Gly Gly

<210> 62  
 <211> 90  
 <212> PRT  
 <213> Danio rerio

<400> 62  
 Met Gly Gly Cys Ser Ala Pro Asn Cys Ser Asn Ser Thr Thr Ile Gly  
 1 5 10 15  
 Lys Gln Leu Phe Arg Phe Pro Lys Asp Pro Val Arg Met Arg Lys Trp  
 20 25 30  
 Leu Val Asn Cys Arg Arg Asp Phe Val Pro Thr Pro Cys Ser Arg Leu  
 35 40 45  
 Cys Gln Asp His Phe Glu Glu Ser Gln Phe Glu Glu Ile Ala Arg Ser  
 50 55 60  
 Pro Ala Gly Gly Arg Lys Leu Lys Pro Asn Ala Ile Pro Thr Leu Phe  
 65 70 75 80  
 Asn Val Pro Asp Pro Pro Ser Pro Val Thr  
 85 90

<210> 63  
 <211> 105  
 <212> PRT  
 <213> Danio rerio

<400> 63  
 Met Val Leu Asn Cys Ala Tyr Pro Gly Cys Leu Asn Leu Phe Lys Lys  
 1 5 10 15  
 Glu Arg Leu Arg Ser Asn Ser Ser Ser His Gly Gly Lys Leu Thr Phe  
 20 25 30  
 His Arg Phe Pro Thr Leu Glu Pro Gly Arg Leu Leu Leu Trp Arg Ala  
 35 40 45  
 Ala Leu Gly Met Asp Pro Asp Thr Pro Met Arg Ser Leu Arg Val Trp  
 50 55 60  
 Arg Ile Cys Ser Glu His Phe Ser Pro Glu Asp Phe Arg Ala Val Asn  
 65 70 75 80  
 Gly Asn Lys Val Leu Leu Lys Ala Ser Ala Val Pro Arg Val Tyr Ser  
 85 90 95  
 Thr Pro Ala Pro Gly Ser Arg Ala Asp  
 100 105

<210> 64  
 <211> 99  
 <212> PRT  
 <213> Danio rerio

<400> 64  
 Met Ala Ser Ser Arg Arg Cys Tyr Cys Ser Val Pro Gly Cys Ser Asn  
 1 5 10 15  
 Ser Lys Lys Arg His Pro Tyr Leu Ser Phe His Asp Phe Pro Lys Asp  
 20 25 30  
 Glu Gly Gln Arg Lys Ser Trp Val Lys Phe Ile Arg Arg Glu Glu Gly  
 35 40 45  
 Pro Phe Phe Gln Ile Lys Arg Gly Ser Thr Phe Val Cys Ser Met His  
 50 55 60  
 Phe Lys Ala Asp Asp Ile Tyr Thr Thr Ile Ser Gly Arg Arg Lys Ile  
 65 70 75 80  
 Asn Pro Gly Ala Ala Pro Arg Leu Phe Ser Trp Asn Asn Trp Ser Thr  
 85 90 95  
 Asp Lys Val

<210> 65  
 <211> 66  
 <212> PRT  
 <213> Danio rerio

<400> 65  
 Phe Pro Lys Glu Asn Val Leu Arg Lys Gln Trp Glu Ile Ala Leu Lys  
 1 5 10 15  
 Arg Lys Gly Phe Ser Ala Ser Glu Ser Ser Val Leu Cys Ser Glu His  
 20 25 30  
 Phe Arg Pro Gln Asp Leu Asp Arg Thr Gly Gln Thr Val Arg Val Arg  
 35 40 45  
 Asp Gly Ala Lys Pro Ser Val Phe Ser Phe Pro Ala His Met Gln Lys  
 50 55 60  
 His Val  
 65

<210> 66  
 <211> 93  
 <212> PRT  
 <213> Danio rerio

<400> 66  
 Ser Ser Glu His Cys Cys Val Pro Leu Cys Gly Ala Ser Ser Arg Phe  
 1 5 10 15  
 Asn Ser Ala Val Ser Phe His Thr Phe Pro Val Ser Thr Glu Ile Arg  
 20 25 30  
 Glu Lys Trp Ile Lys Asn Ile Arg Arg Glu Lys Leu Asn Ile Thr Tyr  
 35 40 45  
 His Thr Arg Val Cys Cys Arg His Phe Thr Thr Asp Asp Leu Ile Gln  
 50 55 60  
 Pro Arg Asn Pro Ile Gly Arg Arg Leu Leu Arg Lys Gly Ala Val Pro  
 65 70 75 80  
 Thr Leu Phe Lys Trp Asn Gly Tyr Ser Asp Ala Glu Ala  
 85 90

<210> 67  
 <211> 93  
 <212> PRT  
 <213> Danio rerio

<400> 67  
 Met Pro Asp Phe Cys Ala Ala Tyr Gly Cys Ser Asn Glu Arg Thr Lys  
 1 5 10 15  
 Lys Leu Lys Asp Lys Gly Ile Thr Phe His Arg Phe Pro Arg Asp Val  
 20 25 30  
 Lys Arg Arg Gln Ala Trp Thr Leu Ala Leu Arg Arg Asp Lys Phe Glu  
 35 40 45  
 Pro Lys Pro Arg Ser Leu Leu Cys Ser Cys His Phe Arg Pro Glu Asp  
 50 55 60  
 Phe Asp Arg Thr Gly Gln Thr Val Arg Leu Arg Asp Gly Val Ile Pro  
 65 70 75 80  
 Ser Ile Phe Asn Phe Ser Asn Pro Leu Ser Lys Leu Ser  
 85 90

<210> 68  
 <211> 97  
 <212> PRT  
 <213> Danio rerio

<400> 68  
 Met Pro Val Cys Ser Ala Tyr Lys Cys Lys Lys Arg Ser Asp Arg Glu  
 1 5 10 15  
 Tyr Lys Glu Ala Tyr Lys Arg Gly Glu Phe Ser Phe His Lys Phe Pro  
 20 25 30  
 Leu Glu Asp Gly Leu Arg Val Arg Glu Trp Leu Arg Arg Met Arg Trp  
 35 40 45  
 Gln Asn Trp Trp Pro Thr Gly Asn Ser Val Leu Cys Ser Asp His Phe  
 50 55 60  
 Glu Lys Asp Cys Phe Glu Gln Val Gly Ser His Lys Arg Leu Arg Lys  
 65 70 75 80  
 Ser Ala Val Pro Thr Ile Phe Asn Phe Pro Lys His Leu Gln Trp Lys  
 85 90 95  
 Val

<210> 69  
 <211> 90  
 <212> PRT  
 <213> Danio rerio

<400> 69  
 Met Val Leu Val Cys Ser Ala Tyr Asn Cys Lys Asn Thr Leu Arg Asn  
 1 5 10 15  
 Lys Ser Val Ser Phe His Leu Phe Pro Leu Lys Asp Pro Ser Leu Leu  
 20 25 30  
 Lys Lys Trp Leu Lys Asn Leu Arg Trp Lys Asp Trp Lys Pro Asn Pro  
 35 40 45  
 Asn Ser Lys Ile Cys Ser Ala His Phe Glu Glu Lys Cys Phe Ile Leu  
 50 55 60  
 Glu Gly Lys Lys Thr Arg Leu His Thr Trp Ala Val Pro Thr Ile Phe  
 65 70 75 80  
 Ser Phe Pro Asn Arg Phe Ser Glu Arg Asn  
 85 90

<210> 70  
 <211> 107  
 <212> PRT  
 <213> Danio rerio

<400> 70  
 Met Asn Ser Ile Ser Leu Lys Tyr Leu Arg Arg Glu Cys Ala Tyr Ser  
 1 5 10 15  
 Arg Tyr Cys Cys Val Pro Phe Cys Lys Ile Ser Ser Arg Phe Asn Ser  
 20 25 30  
 Val Ile Ser Phe His Lys Leu Pro Leu Asp Arg Ala Thr Arg Lys Met  
 35 40 45  
 Trp Leu His Asn Ile Arg Arg Lys Thr Phe Glu Val Ser Pro His Val  
 50 55 60  
 Arg Val Cys Ser Arg His Phe Thr Asn Asp Asp Phe Ile Glu Pro Ser  
 65 70 75 80  
 Tyr Pro Thr Ala Arg Arg Leu Leu Lys Lys Gly Ala Val Pro Thr Leu  
 85 90 95  
 Phe Arg Trp Asn Asn Asp Ser Thr Ser Gly Gln  
 100 105

<210> 71  
 <211> 89  
 <212> PRT  
 <213> Danio rerio

<400> 71  
 Leu Arg Leu Arg Gln Ser Ala Ser Ser His Glu Glu Ser Leu Thr Phe  
 1 5 10 15  
 Tyr Ser Leu Pro Leu Gln Asp Phe Lys Arg Leu Asn Leu Trp Leu Asn  
 20 25 30  
 Ala Val Arg Arg Asp Thr Lys Ser Ser Ile Arg Asn Ile Arg Gly Leu  
 35 40 45  
 Arg Val Cys Ser Glu His Phe Ala Gln Asp Asp Phe Ser Leu Asn Arg  
 50 55 60  
 Gly Ser Lys Arg Arg Leu Lys Ser Thr Ala Val Pro Lys Cys Asn Glu  
 65 70 75 80  
 Ala Leu Pro Gln Ile Arg Arg Ala Gly  
 85

<210> 72  
 <211> 105  
 <212> PRT  
 <213> Danio rerio

<400> 72  
 Met Val Ile Thr Cys Ala Cys Pro Gly Cys Asp Asn Arg Tyr Lys Thr  
 1 5 10 15  
 Leu Arg Leu Arg Ser Asp Ser Lys Phe His Pro Gly Lys Leu Thr Phe  
 20 25 30  
 His Lys Phe Pro Thr Ser Asp Pro Glu Arg Leu Lys Leu Trp Leu Leu  
 35 40 45  
 Ala Leu Gly Leu Asp Ile Asn Thr Pro Leu Ser Val Leu Glu Thr Arg  
 50 55 60  
 Arg Ile Cys Ser Asp His Phe Ser Pro Phe Asp Phe Lys Asp Thr Lys  
 65 70 75 80  
 Gly Ser Ile Val Gln Leu Lys Ser Trp Ala Val Pro Met Asn Leu Ser

85 90 95  
 Glu Gln Phe Val Asp Asp Pro Ser Lys  
 100 105

<210> 73  
 <211> 96  
 <212> PRT  
 <213> Danio rerio

<400> 73  
 Met Pro Asp Cys Cys Ala Ala Ala Asn Cys Lys Gln Ser Thr Asp Gln  
 1 5 10 15  
 Ser Ser Val Ser Phe Phe Glu Phe Pro Leu Asp Pro Asp Arg Cys Arg  
 20 25 30  
 Gln Trp Val Gly Arg Cys Asn Arg Pro Asp Leu Gln Thr Lys Thr Pro  
 35 40 45  
 Glu Asp Leu His Lys Asn Tyr Lys Val Cys Ser Arg His Phe Glu Thr  
 50 55 60  
 Ser Met Ile Cys Gln Gln Ser Ala Val Lys Cys Ile Leu Lys Asp Asp  
 65 70 75 80  
 Ala Val Pro Thr Leu Phe Asn Phe Ser Thr Asn Gln Asp Asn Ala Gln  
 85 90 95

<210> 74  
 <211> 91  
 <212> PRT  
 <213> Danio rerio

<400> 74  
 Met Val Lys Cys Thr Val Gln Gly Cys Ile Asn Phe Ser Asp Leu Arg  
 1 5 10 15  
 Pro Glu Glu Gln Pro Asn Arg Pro Arg Lys Arg Phe Phe Arg Phe Pro  
 20 25 30  
 Lys Asp Lys Val Leu Val Lys Val Trp Leu Ala Ala Leu Arg Asp Thr  
 35 40 45  
 Glu Arg Glu Ile Thr Asp Leu His Arg Ile Cys Glu Asp His Phe Leu  
 50 55 60  
 Ser His His Ile Thr Ala Asp Gly Ile Ser Pro Asp Ala Ile Pro Ile  
 65 70 75 80  
 Met Pro Pro Leu Asp Gly Pro Val Gly Asn Trp  
 85 90

<210> 75  
 <211> 84  
 <212> PRT  
 <213> Danio rerio

<400> 75  
 Met Pro Ile Ser Cys Ser Ala Val Asp Cys Ser Asn Arg Phe Val Lys  
 1 5 10 15  
 Gly Ser Glu Ile Arg Phe Tyr Arg Phe Pro Ile Ser Lys Pro Gln Leu  
 20 25 30  
 Ala Glu Gln Trp Val Arg Ser Leu Gly Arg Lys Asn Phe Val Pro Thr  
 35 40 45  
 Gln Asn Ser Cys Leu Cys Ser Glu His Phe Gln Pro Asp Cys Phe Arg  
 50 55 60  
 Asp Tyr Asn Gly Lys Leu Phe Leu Arg Glu Asp Ala Val Pro Thr Ile  
 65 70 75 80

Phe Ser Asn Ser

<210> 76  
 <211> 95  
 <212> PRT  
 <213> Oryzias latipes

<400> 76  
 Met Pro Asn Phe Cys Ala Ala Pro Asn Cys Thr Arg Lys Ser Thr Gln  
 1 5 10 15  
 Ser Asp Leu Ala Phe Phe Arg Phe Pro Arg Asp Pro Glu Arg Cys Arg  
 20 25 30  
 Ile Trp Val Glu Asn Cys Arg Arg Ala Asp Leu Glu Ala Lys Thr Ala  
 35 40 45  
 Asp Gln Leu Asn Lys His Tyr Arg Leu Cys Ala Lys His Phe Asp Pro  
 50 55 60  
 Ala Met Val Cys Lys Thr Ser Pro Tyr Arg Thr Val Leu Lys Asp Thr  
 65 70 75 80  
 Ala Ile Pro Thr Ile Phe Asp Leu Thr Ser His Leu Lys Asn Pro  
 85 90 95

<210> 77  
 <211> 90  
 <212> PRT  
 <213> Oryzias latipes

<400> 77  
 Met Pro Thr Gly Cys Ala His Ala Asn Cys Arg Asn Val Val Gly Lys  
 1 5 10 15  
 Phe Arg Gly Val Thr Phe His Lys Phe Pro Arg Asp Pro Glu Lys Leu  
 20 25 30  
 Ser Arg Trp Thr Lys Phe Met Lys Arg His Glu Ser Trp Val Pro Lys  
 35 40 45  
 Tyr Tyr Asp Arg Val Cys Ser Val His Phe Ser Ser Glu His Phe Asp  
 50 55 60  
 Arg Thr Gly Gln Thr Val Arg Leu Arg Asp Asn Ala Glu Pro Ser Leu  
 65 70 75 80  
 Pro His Leu Pro Trp Arg Phe Pro Lys Ser  
 85 90

<210> 78  
 <211> 94  
 <212> PRT  
 <213> Oryzias latipes

<400> 78  
 Met Gln Asn Arg Cys Ala Val Leu Thr Cys Pro Ser Gly Lys Thr Asp  
 1 5 10 15  
 Phe Gln Pro Met Phe Arg Phe Pro His Asp Gln Glu Arg Ser Arg Arg  
 20 25 30  
 Trp Val Glu Lys Cys Gln Gly Glu Asn Leu Ile Gly Lys Ser Pro Glu  
 35 40 45  
 Gln Leu Tyr Arg Tyr Tyr Arg Ile Cys Lys Arg His Phe Glu Thr Ser  
 50 55 60  
 Ala Phe Asp Cys Asp Ala Asp Gly Ala Val Leu Lys Lys Asp Ala Val  
 65 70 75 80  
 Pro Thr Ile Phe Asp Ala Ser Val Pro Pro Gln Ser Ser Gln

85

90

<210> 79  
 <211> 92  
 <212> PRT  
 <213> Drosophila melanogaster

<400> 79  
 Met Pro Ala His Cys Ala Val Ile Asn Cys Ser His Lys Tyr Val His  
 1 5 10 15  
 Ala Gly Ser Ile Ser Phe His Arg Phe Pro Phe Lys Arg Lys Asp Leu  
 20 25 30  
 Leu Gln Lys Trp Lys Glu Phe Thr Gln Arg Ser Ala Gln Trp Met Pro  
 35 40 45  
 Ser Lys Trp Ser Ala Leu Cys Ser Arg His Phe Gly Asp Glu Asp Phe  
 50 55 60  
 Asn Cys Ser Asn Asn Arg Lys Thr Leu Lys Lys Asn Ala Val Pro Ser  
 65 70 75 80  
 Ile Arg Val Ser Glu Asp Asp Ser Met Ser Gly His  
 85 90

<210> 80  
 <211> 90  
 <212> PRT  
 <213> Drosophila melanogaster

<400> 80  
 Met Pro Thr Ile Arg Arg Cys Cys Ile Ile Gly Cys Leu Ser Asn Ser  
 1 5 10 15  
 Arg Gln His Pro Ser Met Gln Phe Phe Ala Phe Pro Arg Pro Glu Asn  
 20 25 30  
 Pro Phe His Lys Leu Trp Lys Glu Ala Cys His Ala Ser Leu Arg Arg  
 35 40 45  
 Ile Val Pro Phe Lys Lys Pro Val Val Cys Ala Leu His Phe Asp Pro  
 50 55 60  
 Ser Val Leu Gly Gly Arg Arg Leu Gln Ser Asn Ala Leu Pro Thr Leu  
 65 70 75 80  
 Arg Leu Glu Val Pro Ser Asn Leu Glu Ala  
 85 90

<210> 81  
 <211> 104  
 <212> PRT  
 <213> Drosophila melanogaster

<400> 81  
 Met Arg Cys Ala Val Pro Asn Cys Arg Asn Phe Ser Asp Cys Arg Ser  
 1 5 10 15  
 Lys Arg Asn Ala Ala Gln Gln Gln Arg Leu Gly Phe Phe Arg Phe Pro  
 20 25 30  
 Lys Cys Pro Asp Thr Phe Lys Ala Trp Leu Ala Phe Cys Gly Tyr Thr  
 35 40 45  
 Glu Glu Ser Leu Lys Leu Lys Asn Pro Cys Ile Cys Ile Glu His Phe  
 50 55 60  
 Lys Asp Glu Asp Ile Glu Gly Ser Leu Lys Phe Glu Met Gly Leu Ala  
 65 70 75 80  
 Lys Lys Arg Thr Leu Arg Pro Gly Ala Val Pro Cys Val Asn Lys Ser  
 85 90 95



Gln Glu Ser Gly Ser Asp Arg Ala  
100

<210> 82

<211> 96

<212> PRT

<213> *Drosophila melanogaster*

<400> 82

Met	Gly	Gly	Thr	Lys	Cys	Cys	Phe	Arg	Asp	Cys	Pro	Val	Gly	Ser	Ser
1				5					10					15	
Arg	Asn	Pro	Asn	Met	His	Phe	Phe	Lys	Phe	Pro	Val	Lys	Asp	Pro	Lys
			20					25					30		
Arg	Leu	Lys	Asp	Trp	Val	Arg	Asn	Cys	Ser	Asn	Pro	Asp	Val	Ser	Asn
		35					40					45			
Ala	Pro	Pro	Ser	Lys	Leu	Ala	Ala	Lys	Thr	Val	Cys	Ala	Arg	His	Phe
	50					55					60				
Arg	Ala	Glu	Cys	Phe	Met	Asn	Tyr	Lys	Met	Asp	Arg	Leu	Ile	Pro	Met
65					70				75					80	
Gln	Thr	Pro	Thr	Leu	Phe	Arg	Ile	Asn	Arg	Asp	Leu	Ala	Leu	Asp	Tyr
				85					90					95	

<210> 83

<211> 96

<212> PRT

<213> *Drosophila melanogaster*

<400> 83

Met	Ala	Thr	Arg	Ser	Cys	Ala	Tyr	Lys	Asp	Cys	Glu	Tyr	Tyr	Tyr	Val
1				5					10					15	
Gly	His	Glu	Asn	Ala	Leu	Thr	Lys	Gly	Arg	Thr	Leu	Phe	Ala	Phe	Pro
			20					25					30		
Lys	Gln	Pro	Gln	Arg	Ala	Arg	Ile	Trp	His	Glu	Asn	Gly	Gln	Val	His
		35					40					45			
Pro	Lys	Ile	Pro	His	Ser	Gln	Leu	Phe	Met	Cys	Ser	Leu	His	Phe	Asp
	50					55				60					
Arg	Lys	Phe	Ile	Ser	Ser	Ser	Lys	Asn	Arg	Thr	Leu	Leu	Val	Gly	Glu
65					70				75					80	
Ala	Val	Pro	Phe	Pro	Tyr	Glu	Glu	Ser	Ser	Ser	Lys	Pro	Glu	Glu	Glu
				85					90					95	

<210> 84

<211> 87

<212> PRT

<213> *Drosophila melanogaster*

<400> 84

Met	Lys	Tyr	Cys	Lys	Phe	Cys	Cys	Lys	Ala	Val	Thr	Gly	Val	Lys	Leu
1				5					10					15	
Ile	His	Val	Pro	Lys	Cys	Ala	Ile	Lys	Arg	Lys	Leu	Trp	Glu	Gln	Ser
			20					25					30		
Leu	Gly	Cys	Ser	Leu	Gly	Glu	Asn	Ser	Gln	Ile	Cys	Asp	Thr	His	Phe
		35					40					45			
Asn	Asp	Ser	Gln	Trp	Lys	Ala	Ala	Pro	Ala	Lys	Gly	Gln	Thr	Phe	Lys
	50					55					60				
Arg	Arg	Arg	Leu	Asn	Ala	Asp	Ala	Val	Pro	Ser	Lys	Val	Ile	Glu	Pro
65					70				75					80	
Glu	Pro	Glu	Lys	Ile	Lys	Glu									

85

<210> 85  
 <211> 92  
 <212> PRT  
 <213> Anopheles gambiae

<400> 85  
 Met Pro Ala Ser Cys Val Ile Pro Asp Cys Asp Leu Lys Tyr Thr His  
 1 5 10 15  
 Gly Asp Asp Val Ser Phe His Lys Phe Pro Leu Lys Ser Pro Glu Leu  
 20 25 30  
 Leu Lys Gln Trp Ile Gln Phe Thr Gly Arg Asp Glu Gly Trp His Pro  
 35 40 45  
 Thr Lys Trp Ser Ala Leu Cys Ser Arg His Phe Val Ala Ser Asp Phe  
 50 55 60  
 Lys Gly Cys Ala Ala Arg Lys Ile Leu Leu Pro Thr Ala Val Pro Ser  
 65 70 75 80  
 Val Arg Asn Ala Val Ala Ala Lys Ala Gln Pro Asn  
 85 90

<210> 86  
 <211> 108  
 <212> PRT  
 <213> Anopheles gambiae

<400> 86  
 Met Ser Ala Val Arg Ser Cys Ala Leu Cys Gln Asn Arg Ser Asn Ile  
 1 5 10 15  
 Thr Asp Gln Gln Thr Asp Asp Ala Leu Glu Arg Ile Thr Tyr His Lys  
 20 25 30  
 Phe Pro Thr Asn Pro Val Arg Arg Asp Arg Trp Ile Glu Phe Cys Asp  
 35 40 45  
 Leu Pro Lys Glu Ser Phe Pro Lys Ser Ala Tyr Lys Phe Leu Cys Ser  
 50 55 60  
 Ser His Phe Thr Pro Glu Cys Phe Glu Arg Asp Leu Arg Gly Glu Leu  
 65 70 75 80  
 Leu Tyr Gly Thr Lys Arg Met Thr Leu Gln Lys Asp Ala Met Pro Thr  
 85 90 95  
 Ile Arg Ser Val Ser Gln Gln Leu Lys Arg Thr Thr  
 100 105

<210> 87  
 <211> 100  
 <212> PRT  
 <213> Anopheles gambiae

<400> 87  
 Met Trp Asp Cys Ala Val Ile Gly Cys Pro Asn Ser Arg Phe Asn Ala  
 1 5 10 15  
 Gln Lys Thr Arg Pro Arg Ile Ser Phe His Val Phe Pro His Pro Val  
 20 25 30  
 Arg Glu Ser Asn Arg Phe Arg Arg Trp Leu Ala Leu Ile Asn Asn Pro  
 35 40 45  
 Arg Leu Phe Arg Leu Asp Pro Leu Asn Val Phe Lys Ser Val Arg Val  
 50 55 60  
 Cys Arg Arg His Phe Gly Pro Asp Cys Phe Asn Gly Val Cys Arg Asn  
 65 70 75 80

Leu Leu Pro Thr Ala Ile Pro Thr Leu Asn Leu Pro Glu Val Arg Pro  
                             85                            90                            95  
 Val Ala Leu Val  
                             100

<210> 88  
 <211> 95  
 <212> PRT  
 <213> Anopheles gambiae

<400> 88  
 Met Gly Ile Arg Lys Cys Ile Val Pro Glu Cys Pro Ser Ser Ser Ala  
   1                            5                            10                            15  
 Arg Pro Glu Asp Arg Gly Val Thr Tyr His Lys Ile Pro Tyr Leu Asp  
                             20                            25                            30  
 Glu Met Lys Arg Leu Trp Ile Val Ala Cys His Leu Pro Asp Asp Tyr  
                             35                            40                            45  
 Phe Ala Thr Lys Ala Ser Asn Val Cys Ser Arg His Phe Arg Arg Ala  
                             50                            55                            60  
 Asp Phe Gln Glu Phe Lys Gly Lys Lys Tyr Val Leu Lys Leu Gly Val  
  65                            70                            75                            80  
 Val Pro Thr Val Phe Pro Trp Thr Val Thr Lys Pro Pro Gly Glu  
                             85                            90                            95

<210> 89  
 <211> 107  
 <212> PRT  
 <213> Anopheles gambiae

<400> 89  
 Met Gly Lys Ile Ser Gly Ser His Cys Leu Val Leu Gly Cys Arg Asn  
   1                            5                            10                            15  
 Arg Gln Leu Leu Asn Gln Ala Asn Ile Arg Ser Tyr Phe Arg Phe Pro  
                             20                            25                            30  
 Arg Asp Ala Asp Leu Cys Lys Lys Trp Val Asp Phe Cys Asn Arg Pro  
                             35                            40                            45  
 Glu Leu Tyr Lys Lys Tyr Asp Glu Asn Gly Pro Glu Tyr Leu Tyr Lys  
                             50                            55                            60  
 Ser Ser Arg Ile Cys Ser Asp His Phe Gln Pro Ala Asp Phe Asn Asn  
  65                            70                            75                            80  
 Pro Asn Leu Phe Ser Gln Gly Leu Lys Lys Gly Ser Val Pro Ser Val  
                             85                            90                            95  
 Asn Pro Ala Asn Leu Glu Ala Ala Lys Pro His  
                             100                            105

<210> 90  
 <211> 104  
 <212> PRT  
 <213> Anopheles gambiae

<400> 90  
 Met Thr Asn Cys Ser Cys Ala Val Ala Asp Cys Asn Asn Asn Arg Arg  
   1                            5                            10                            15  
 Asn Val Arg Lys Arg Met Leu Asp Ile Gly Phe His Thr Phe Pro Ser  
                             20                            25                            30  
 Asp Pro Val Gln Arg Gln Arg Trp Val Lys Phe Cys Gln Arg Glu Pro  
                             35                            40                            45  
 Ser Trp Gln Pro Lys Ser Cys Asp Ser Met Cys Ser Val His Phe Lys

50                      55                      60  
 Asp Thr Asp Tyr Gln Met Ser His Ser Pro Leu Ile Arg Leu Ala Thr  
 65                      70                      75                      80  
 Asn Leu Arg Arg Leu Lys Pro Asp Val Ile Pro Thr Ile Arg Lys Gly  
                     85                      90                      95  
 Arg Ala Ile Pro Val Ala Ala Arg  
                     100

<210> 91  
 <211> 95  
 <212> PRT  
 <213> Anopheles gambiae

<400> 91  
 Met Gly Gly Cys Arg Cys Thr Phe Arg Asp Cys Glu Asn Gly Thr Ala  
 1                      5                      10                      15  
 Ser Arg Lys Glu Leu His Tyr Phe Arg Tyr Pro Val Arg Asp Gln Glu  
                     20                      25                      30  
 Arg Leu Ile Glu Trp Ala Lys Asn Ala Asp Arg Leu Glu Phe Val Asp  
                     35                      40                      45  
 Leu Pro Val Asp Lys Val Ser Asn Lys Val Val Cys Gln Glu His Phe  
                     50                      55                      60  
 Glu Arg Lys Met Phe Met Asn Asp Leu Arg Asp Arg Leu Thr Lys Met  
 65                      70                      75                      80  
 Ala Ile Pro Arg Leu Met Val Met Pro Asp Glu Thr Ile Val Asn  
                     85                      90                      95

<210> 92  
 <211> 97  
 <212> PRT  
 <213> Anopheles gambiae

<400> 92  
 Met Lys Cys Phe Val Ser Gly Cys Asp Thr Asp Asp Asn Val Val Ser  
 1                      5                      10                      15  
 Tyr Thr Ser Val Phe Tyr Val Asn Cys Pro Thr Asp Pro Thr Ile Gln  
                     20                      25                      30  
 Gln Gln Trp Phe Thr Leu Leu Glu Val Thr Asp Pro Asp Ala Met Arg  
                     35                      40                      45  
 Ala Leu Val Asp Gly Arg Ser Lys Val Cys Ser Cys His Phe Thr Glu  
                     50                      55                      60  
 Asp Cys Phe Gly His His Pro Val Tyr Gly Tyr Arg Tyr Leu Leu Ala  
 65                      70                      75                      80  
 Thr Ala Leu Pro Thr Val Phe Pro Pro Arg Lys Glu Ile Glu Gln Pro  
                     85                      90                      95  
 Lys

<210> 93  
 <211> 92  
 <212> PRT  
 <213> Bombyx mori

<400> 93  
 Met Pro Arg Cys Ser Val Ile Val Cys Lys Asn Asn Ser Cys Ile Val  
 1                      5                      10                      15  
 Asn Tyr Lys Lys Asp Ser Ile Ser Phe His Thr Tyr Pro Lys Asp Pro  
                     20                      25                      30

Lys Ile Lys Glu Met Trp Ile Asn Ala Thr Gly Arg Gly Pro Ser Trp  
           35                          40                          45  
 Phe Pro Thr Lys Asn His Thr Ile Cys Ser Ser His Phe Glu Pro Lys  
       50                          55                          60  
 Cys Phe Gln Pro Leu Lys Lys Val Arg Arg Leu Phe Glu Trp Ser Val  
       65                          70                          75                          80  
 Pro Thr Leu Lys Leu Arg Met Val Leu Met Asn Tyr  
                           85                          90

<210> 94  
 <211> 96  
 <212> PRT  
 <213> Bombyx mori

<400> 94  
 Met Pro Asp Thr His Arg Thr Cys Glu Val Cys Gly Ile Lys Glu Arg  
   1                          5                          10                          15  
 His Leu Thr Glu Lys Arg Phe Phe Ala Arg Phe Pro Leu Asp Val Asn  
                           20                          25                          30  
 Arg Cys Lys Gln Trp Val Lys Met Val Gly Lys Glu Asp Leu Ala Tyr  
       35                          40                          45  
 Leu Gln Val His Met Leu His Asp Leu Lys His Val Cys Glu Ala His  
       50                          55                          60  
 Phe Ser Arg Arg Asp Phe Thr Lys Ser Lys Lys Arg Leu Lys Lys Arg  
       65                          70                          75                          80  
 Ala Val Pro Lys Leu Asn Leu Thr Leu Pro Pro Leu Arg Asp Glu Ile  
                           85                          90                          95

<210> 95  
 <211> 89  
 <212> PRT  
 <213> Caenorhabditis elegans

<400> 95  
 Met Pro Thr Thr Cys Gly Phe Pro Asn Cys Lys Phe Arg Ser Arg Tyr  
   1                          5                          10                          15  
 Arg Gly Leu Glu Asp Asn Arg His Phe Tyr Arg Ile Pro Lys Arg Pro  
                           20                          25                          30  
 Leu Ile Leu Arg Gln Arg Trp Leu Thr Ala Ile Gly Arg Thr Glu Glu  
       35                          40                          45  
 Thr Val Val Ser Gln Leu Arg Ile Cys Ser Ala His Phe Glu Gly Gly  
       50                          55                          60  
 Glu Lys Lys Glu Gly Asp Ile Pro Val Pro Asp Pro Thr Val Asp Lys  
       65                          70                          75                          80  
 Gln Ile Lys Ile Glu Leu Pro Pro Lys  
                           85

<210> 96  
 <211> 100  
 <212> PRT  
 <213> Caenorhabditis elegans

<400> 96  
 Met Tyr Gly Val Gln Ser Glu Cys Val Leu Cys Ala His Ala Asn Asp  
   1                          5                          10                          15  
 Cys Thr Ala Met Ile Pro Phe Pro Gly Pro Asp Asp Glu Lys Leu Arg  
                           20                          25                          30  
 Thr Lys Trp Ile Asn Ser Met Cys Arg Glu Pro Trp Ile Tyr Arg Tyr

35 40 45  
 Leu Ser Thr Arg Leu Glu Lys Pro Gly Arg His Tyr Leu Cys Ala Ser  
 50 55 60  
 His Phe Asn Arg Asn Ser Leu Arg Tyr His Ala Gly Leu Gly Leu Trp  
 65 70 75 80  
 Arg Arg Ala Ala Ala Cys Pro Val Leu Ala Cys Thr Thr Asp Glu Glu  
 85 90 95  
 Arg Gln Glu Val  
 100

<210> 97  
 <211> 86  
 <212> PRT  
 <213> Caenorhabditis elegans

<400> 97  
 Met Glu His Pro Leu Gln Cys Cys Tyr Cys Leu Glu Val Tyr Glu Lys  
 1 5 10 15  
 Arg Tyr Met Thr Gln Val Pro Lys Thr Glu Gln Arg Ile Ala Arg Trp  
 20 25 30  
 Val Ala Ile Leu Gly Glu Gln Phe Arg Ile Arg Leu Arg Met Lys Pro  
 35 40 45  
 Ala Asn Tyr Met Cys Arg Lys His Phe Pro Gln Ala Asp Phe Ser Ser  
 50 55 60  
 Arg Gly Arg Leu Leu Lys Thr Ala Val Pro Asn Val Val Ser Gln Glu  
 65 70 75 80  
 Lys Val Leu Ala Phe Lys  
 85

<210> 98  
 <211> 97  
 <212> PRT  
 <213> Caenorhabditis elegans

<400> 98  
 Asn Leu Thr His Lys Pro Cys Thr Val Cys Asn Arg Val Met Lys Ser  
 1 5 10 15  
 Gly Glu Met His Leu Asn Phe Pro Ala Asp Leu Asp Arg Arg Arg Ile  
 20 25 30  
 Trp Ala Asn Leu Leu Gly Phe Lys Tyr Lys Asp Ile Leu Arg Ser Lys  
 35 40 45  
 Met Gly Pro Val Ser Phe Ser Ile Ala Ala Gly Pro Ile Cys Thr Glu  
 50 55 60  
 His Phe Ala Glu Glu Cys Phe Arg Asn His Asn Phe Asn Lys Ser Ala  
 65 70 75 80  
 Ile Glu Ala Phe Gly Val Pro Val Ala Ile Ser Pro Asp Val Lys Thr  
 85 90 95  
 Thr

<210> 99  
 <211> 210  
 <212> PRT  
 <213> Mus musculus

<400> 99  
 Met Val Gln Ser Cys Ser Ala Tyr Gly Cys Lys Asn Arg Tyr Asp Lys  
 1 5 10 15

Asp Lys Pro Val Ser Phe His Lys Phe Pro Leu Thr Arg Pro Ser Leu  
 20 25 30  
 Cys Lys Gln Trp Glu Ala Ala Val Lys Arg Lys Asn Phe Lys Pro Thr  
 35 40 45  
 Lys Tyr Ser Ser Ile Cys Ser Glu His Phe Thr Pro Asp Cys Phe Lys  
 50 55 60  
 Arg Glu Cys Asn Asn Lys Leu Leu Lys Glu Asn Ala Val Pro Thr Ile  
 65 70 75 80  
 Phe Leu Tyr Ile Glu Pro His Glu Lys Lys Glu Asp Leu Glu Ser Gln  
 85 90 95  
 Glu Gln Leu Pro Ser Pro Ser Pro Pro Ala Ser Gln Val Asp Ala Ala  
 100 105 110  
 Ile Gly Leu Leu Met Pro Pro Leu Gln Thr Pro Asp Asn Leu Ser Val  
 115 120 125  
 Phe Cys Asp His Asn Tyr Thr Val Glu Asp Thr Met His Gln Arg Lys  
 130 135 140  
 Arg Ile Leu Gln Leu Glu Gln Gln Val Glu Lys Leu Arg Lys Lys Leu  
 145 150 155 160  
 Lys Thr Ala Gln Gln Arg Cys Arg Arg Gln Glu Arg Gln Leu Glu Lys  
 165 170 175  
 Leu Lys Glu Val Val His Phe Gln Arg Glu Lys Asp Asp Ala Ser Glu  
 180 185 190  
 Arg Gly Tyr Val Ile Leu Pro Asn Asp Tyr Phe Glu Ile Val Glu Val  
 195 200 205  
 Pro Ala  
 210

&lt;210&gt; 100

&lt;211&gt; 217

&lt;212&gt; PRT

&lt;213&gt; Mus musculus

&lt;400&gt; 100

Met Pro Thr Asn Cys Ala Ala Ala Gly Cys Ala Ala Thr Tyr Asn Lys  
 1 5 10 15  
 His Ile Asn Ile Ser Phe His Arg Phe Pro Leu Asp Pro Lys Arg Arg  
 20 25 30  
 Lys Glu Trp Val Arg Leu Val Arg Arg Lys Asn Phe Val Pro Gly Lys  
 35 40 45  
 His Thr Phe Leu Cys Ser Lys His Phe Glu Ala Ser Cys Phe Asp Leu  
 50 55 60  
 Thr Gly Gln Thr Arg Arg Leu Lys Met Asp Ala Val Pro Thr Ile Phe  
 65 70 75 80  
 Asp Phe Cys Thr His Ile Lys Ser Leu Lys Leu Lys Ser Arg Asn Leu  
 85 90 95  
 Leu Lys Thr Asn Asn Ser Phe Pro Pro Thr Gly Pro Cys Asn Leu Lys  
 100 105 110  
 Leu Asn Gly Ser Gln Gln Val Leu Glu His Ser Tyr Ala Phe Arg  
 115 120 125  
 Asn Pro Met Glu Ala Lys Lys Arg Ile Ile Lys Leu Glu Lys Glu Ile  
 130 135 140  
 Ala Ser Leu Arg Lys Lys Met Lys Thr Cys Leu Gln Arg Glu Arg Arg  
 145 150 155 160  
 Ala Thr Arg Arg Trp Ile Lys Ala Thr Cys Phe Val Lys Ser Leu Glu  
 165 170 175  
 Ala Ser Asn Met Leu Pro Lys Gly Ile Ser Glu Gln Ile Leu Pro Thr  
 180 185 190  
 Ala Leu Ser Asn Leu Pro Leu Glu Asp Leu Lys Ser Leu Glu Gln Asp  
 195 200 205  
 Gln Gln Asp Lys Thr Val Pro Ile Leu

210

215

<210> 101  
 <211> 218  
 <212> PRT  
 <213> Mus musculus

<400> 101  
 Met Pro Lys Ser Cys Ala Ala Arg Gln Cys Cys Asn Arg Tyr Ser Ser  
 1 5 10 15  
 Arg Arg Lys Gln Leu Thr Phe His Arg Phe Pro Phe Ser Arg Pro Glu  
 20 25 30  
 Leu Leu Arg Glu Trp Val Leu Asn Ile Gly Arg Ala Asp Phe Lys Pro  
 35 40 45  
 Lys Gln His Thr Val Ile Cys Ser Glu His Phe Arg Pro Glu Cys Phe  
 50 55 60  
 Ser Ala Phe Gly Asn Arg Lys Asn Leu Lys His Asn Ala Val Pro Thr  
 65 70 75 80  
 Val Phe Ala Phe Gln Asn Pro Thr Glu Val Cys Pro Glu Val Gly Ala  
 85 90 95  
 Gly Gly Asp Ser Ser Gly Arg Asn Met Asp Thr Thr Leu Glu Glu Leu  
 100 105 110  
 Gln Pro Pro Thr Pro Glu Gly Pro Val Gln Gln Val Leu Pro Asp Arg  
 115 120 125  
 Glu Ala Met Glu Ala Thr Glu Ala Ala Gly Leu Pro Ala Ser Pro Leu  
 130 135 140  
 Gly Leu Lys Arg Pro Leu Pro Gly Gln Pro Ser Asp His Ser Tyr Ala  
 145 150 155 160  
 Leu Ser Asp Leu Asp Thr Leu Lys Lys Lys Leu Phe Leu Thr Leu Lys  
 165 170 175  
 Glu Asn Lys Arg Leu Arg Lys Arg Leu Lys Ala Gln Arg Leu Leu Leu  
 180 185 190  
 Arg Arg Thr Cys Gly Arg Leu Arg Ala Tyr Arg Glu Gly Gln Pro Gly  
 195 200 205  
 Pro Arg Ala Arg Arg Pro Ala Gln Gly Ser  
 210 215

<210> 102  
 <211> 205  
 <212> PRT  
 <213> Mus musculus

<400> 102  
 Met Val Ile Cys Cys Ala Ala Val Asn Cys Ser Asn Arg Gln Gly Lys  
 1 5 10 15  
 Gly Glu Lys Arg Ala Val Ser Phe His Arg Phe Pro Leu Lys Asp Ser  
 20 25 30  
 Lys Arg Leu Ile Gln Trp Leu Lys Ala Val Gln Arg Asp Asn Trp Thr  
 35 40 45  
 Pro Thr Lys Tyr Ser Phe Leu Cys Ser Glu His Phe Thr Lys Asp Ser  
 50 55 60  
 Phe Ser Lys Arg Leu Glu Asp Gln His Arg Leu Leu Lys Pro Thr Ala  
 65 70 75 80  
 Val Pro Ser Ile Phe His Leu Ser Glu Lys Lys Arg Gly Ala Gly Gly  
 85 90 95  
 His Gly His Ala Arg Arg Lys Thr Thr Ala Ala Met Arg Gly His Thr  
 100 105 110  
 Ser Ala Glu Thr Gly Lys Gly Thr Ile Gly Ser Ser Leu Ser Ser Ser  
 115 120 125



Asp Asn Leu Met Ala Lys Pro Glu Ser Arg Lys Leu Lys Arg Ala Ser  
 130 135 140  
 Leu Gln Asp Asp Ala Ala Pro Lys Val Thr Pro Gly Ala Val Ser Gln  
 145 150 155 160  
 Glu Gln Gly Gln Ser Leu Glu Lys Thr Pro Gly Asp Asp Pro Ala Ala  
 165 170 175  
 Pro Leu Ala Arg Gly Gln Glu Glu Ala Gln Ala Ser Ala Thr Glu Ala  
 180 185 190  
 Asp His Gln Lys Ala Ser Ser Ser Thr Asp Ala Glu Gly  
 195 200 205

<210> 103  
 <211> 186  
 <212> PRT  
 <213> Mus musculus

<400> 103  
 Ile Leu Gln Ala Phe Gly Ser Leu Lys Lys Gly Asp Val Leu Cys Ser  
 1 5 10 15  
 Arg His Phe Lys Lys Thr Asp Phe Asp Arg Ser Thr Leu Asn Thr Lys  
 20 25 30  
 Leu Lys Ala Gly Ala Ile Pro Ser Ile Phe Glu Cys Pro Tyr His Leu  
 35 40 45  
 Gln Glu Lys Arg Glu Lys Leu His Cys Arg Lys Asn Phe Leu Leu Lys  
 50 55 60  
 Thr Leu Pro Ile Thr His His Gly Arg Gln Leu Val Gly Ala Ser Cys  
 65 70 75 80  
 Ile Glu Glu Phe Glu Pro Gln Phe Ile Phe Glu His Ser Tyr Ser Val  
 85 90 95  
 Met Asp Ser Pro Lys Lys Leu Lys His Lys Leu Asp Arg Val Ile Ile  
 100 105 110  
 Glu Leu Glu Asn Thr Lys Glu Ser Leu Arg Asn Val Leu Ala Arg Glu  
 115 120 125  
 Lys His Phe Gln Lys Ser Leu Arg Lys Thr Ile Met Glu Leu Lys Asp  
 130 135 140  
 Glu Ser Leu Ile Ser Gln Glu Thr Ala Asn Ser Leu Gly Ala Phe Cys  
 145 150 155 160  
 Trp Glu Cys Tyr His Glu Ser Thr Ala Gly Gly Cys Ser Cys Glu Val  
 165 170 175  
 Ile Ser Tyr Met Leu His Leu Gln Leu Thr  
 180 185

<210> 104  
 <211> 194  
 <212> PRT  
 <213> Mus musculus

<400> 104  
 Met Pro Arg His Cys Ser Ala Ala Gly Cys Cys Thr Arg Asp Thr Arg  
 1 5 10 15  
 Glu Thr Arg Asn Arg Gly Ile Ser Phe His Arg Leu Pro Lys Lys Asp  
 20 25 30  
 Asn Pro Arg Arg Gly Leu Trp Leu Ala Asn Cys Gln Arg Leu Asp Pro  
 35 40 45  
 Ser Gly Gln Gly Leu Trp Asp Pro Thr Ser Glu Tyr Ile Tyr Phe Cys  
 50 55 60  
 Ser Lys His Phe Glu Glu Asn Cys Phe Glu Leu Val Gly Ile Ser Gly  
 65 70 75 80  
 Tyr His Arg Leu Lys Glu Gly Ala Val Pro Thr Ile Phe Glu Ser Phe

85 90 95  
 Ser Lys Leu Arg Arg Thr Ala Lys Thr Lys Gly His Gly Tyr Pro Pro  
 100 105 110  
 Gly Leu Pro Asp Val Ser Arg Leu Arg Arg Cys Arg Lys Arg Cys Ser  
 115 120 125  
 Glu Arg Gln Gly Pro Thr Thr Pro Phe Ser Pro Pro Arg Ala Asp  
 130 135 140  
 Ile Ile Cys Phe Pro Val Glu Glu Ala Ser Ala Pro Ala Thr Leu Pro  
 145 150 155 160  
 Ala Ser Pro Ala Val Arg Leu Asp Pro Gly Leu Asn Ser Pro Phe Ser  
 165 170 175  
 Asp Leu Leu Gly Pro Leu Gly Ala Gln Ala Asp Glu Ala Gly Cys Ser  
 180 185 190  
 Thr Gln

&lt;210&gt; 105

&lt;211&gt; 305

&lt;212&gt; PRT

&lt;213&gt; Mus musculus

&lt;400&gt; 105

Met Pro Gly Phe Thr Cys Cys Val Pro Gly Cys Tyr Asn Asn Ser His  
 1 5 10 15  
 Arg Asp Lys Ala Leu His Phe Tyr Thr Phe Pro Lys Asp Ala Glu Leu  
 20 25 30  
 Arg Arg Leu Trp Leu Lys Asn Val Ser Arg Ala Gly Val Ser Gly Cys  
 35 40 45  
 Phe Ser Thr Phe Gln Pro Thr Thr Gly His Arg Leu Cys Ser Val His  
 50 55 60  
 Phe Gln Gly Gly Arg Lys Thr Tyr Thr Val Arg Val Pro Thr Ile Phe  
 65 70 75 80  
 Pro Leu Arg Gly Val Asn Glu Arg Lys Val Ala Arg Arg Pro Ala Gly  
 85 90 95  
 Ala Ala Ala Ala Arg Arg Arg Gln Gln Gln Gln Gln Gln Gln  
 100 105 110  
 Gln Gln Gln Gln Gln Gln Leu Gln Gln Gln Pro Ser Pro Ser Ser  
 115 120 125  
 Ser Thr Ala Gln Thr Thr Gln Leu Gln Pro Asn Leu Val Ser Ala Ser  
 130 135 140  
 Ala Ala Val Leu Leu Thr Leu Gln Ala Ala Val Asp Ser Asn Gln Ala  
 145 150 155 160  
 Pro Gly Ser Val Val Pro Val Ser Thr Thr Pro Ser Gly Asp Asp Val  
 165 170 175  
 Lys Pro Ile Asp Leu Thr Val Gln Val Glu Phe Ala Ala Ala Glu Gly  
 180 185 190  
 Ala Ala Ala Ala Ala Ala Ser Glu Leu Glu Ala Ala Thr Ala Gly  
 195 200 205  
 Leu Glu Ala Ala Glu Cys Thr Leu Gly Pro Gln Leu Val Val Val Gly  
 210 215 220  
 Glu Glu Gly Phe Pro Asp Thr Gly Ser Asp His Ser Tyr Ser Leu Ser  
 225 230 235 240  
 Ser Gly Thr Thr Glu Glu Leu Leu Arg Lys Leu Asn Glu Gln Arg  
 245 250 255  
 Asp Ile Leu Ala Leu Met Glu Val Lys Met Lys Glu Met Lys Gly Ser  
 260 265 270  
 Ile Arg His Leu Arg Leu Thr Glu Ala Lys Leu Arg Glu Glu Leu Arg  
 275 280 285  
 Glu Lys Asp Arg Leu Leu Ala Met Ala Val Ile Arg Lys Lys His Gly  
 290 295 300

Met  
305

<210> 106  
<211> 305  
<212> PRT  
<213> Mus musculus

<400> 106

Met	Pro	Gly	Phe	Thr	Cys	Cys	Val	Pro	Gly	Cys	Tyr	Asn	Asn	Ser	His
1				5					10					15	
Arg	Asp	Lys	Ala	Leu	His	Phe	Tyr	Thr	Phe	Pro	Lys	Asp	Ala	Glu	Leu
			20					25					30		
Arg	Arg	Leu	Trp	Leu	Lys	Asn	Val	Ser	Arg	Ala	Gly	Val	Ser	Gly	Cys
		35					40					45			
Phe	Ser	Thr	Phe	Gln	Pro	Thr	Thr	Gly	His	Arg	Leu	Cys	Ser	Val	His
	50					55					60				
Phe	Gln	Gly	Gly	Arg	Lys	Thr	Tyr	Thr	Val	Arg	Val	Pro	Thr	Ile	Phe
65					70					75				80	
Pro	Leu	Arg	Gly	Val	Asn	Glu	Arg	Lys	Val	Ala	Arg	Arg	Pro	Ala	Gly
				85					90					95	
Ala	Ala	Ala	Ala	Arg	Arg	Arg	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln
			100					105					110		
Gln	Gln	Gln	Gln	Gln	Gln	Leu	Gln	Gln	Gln	Gln	Pro	Ser	Pro	Ser	Ser
		115					120					125			
Ser	Thr	Ala	Gln	Thr	Thr	Gln	Leu	Gln	Pro	Asn	Leu	Val	Ser	Ala	Ser
	130					135					140				
Ala	Ala	Val	Leu	Leu	Thr	Leu	Gln	Ala	Ala	Val	Asp	Ser	Asn	Gln	Ala
145					150					155					160
Pro	Gly	Ser	Val	Val	Pro	Val	Ser	Thr	Thr	Pro	Ser	Gly	Asp	Asp	Val
				165					170					175	
Lys	Pro	Ile	Asp	Leu	Thr	Val	Gln	Val	Glu	Phe	Ala	Ala	Ala	Glu	Gly
			180					185					190		
Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ser	Glu	Leu	Glu	Ala	Ala	Thr	Ala	Gly
		195					200					205			
Leu	Glu	Ala	Ala	Glu	Cys	Thr	Leu	Gly	Pro	Gln	Leu	Val	Val	Val	Gly
	210					215					220				
Glu	Glu	Gly	Phe	Pro	Asp	Thr	Gly	Ser	Asp	His	Ser	Tyr	Ser	Leu	Ser
225					230				235					240	
Ser	Gly	Thr	Thr	Glu	Glu	Leu	Leu	Arg	Lys	Leu	Asn	Glu	Gln	Arg	
			245					250					255		
Asp	Ile	Leu	Ala	Leu	Met	Glu	Val	Lys	Met	Lys	Glu	Met	Lys	Gly	Ser
			260					265					270		
Ile	Arg	His	Leu	Arg	Leu	Thr	Glu	Ala	Lys	Leu	Arg	Glu	Glu	Leu	Arg
		275					280					285			
Glu	Lys	Asp	Arg	Leu	Leu	Ala	Met	Ala	Val	Ile	Arg	Lys	Lys	His	Gly
	290					295					300				

Met  
305

<210> 107  
<211> 652  
<212> PRT  
<213> Mus musculus

<400> 107

Met	Pro	Asn	Phe	Cys	Ala	Ala	Pro	Asn	Cys	Thr	Arg	Lys	Ser	Thr	Gln
1				5					10					15	
Ser	Asp	Leu	Ala	Phe	Phe	Arg	Phe	Pro	Arg	Asp	Pro	Ala	Arg	Cys	Gln

20 25 30  
 Lys Trp Val Glu Asn Cys Arg Arg Ala Asp Leu Glu Asp Lys Thr Pro  
 35 40 45  
 Asp Gln Leu Asn Lys His Tyr Arg Leu Cys Ala Lys His Phe Glu Thr  
 50 55 60  
 Ser Met Ile Cys Arg Thr Ser Pro Tyr Arg Thr Val Leu Arg Asp Asn  
 65 70 75 80  
 Ala Ile Pro Thr Ile Phe Asp Leu Thr Ser His Leu Asn Asn Pro His  
 85 90 95  
 Ser Arg His Arg Lys Arg Ile Lys Glu Leu Ser Glu Asp Glu Ile Arg  
 100 105 110  
 Thr Leu Lys Gln Lys Lys Ile Glu Thr Ser Glu Gln Glu Gln Glu  
 115 120 125  
 Thr Asn Thr Asn Ala Gln Asn Pro Ser Ala Glu Ala Val Asn Gln Gln  
 130 135 140  
 Asp Ala Asn Val Leu Pro Leu Thr Leu Glu Glu Lys Glu Asn Lys Glu  
 145 150 155 160  
 Tyr Leu Lys Ser Leu Phe Glu Ile Leu Val Leu Met Gly Lys Gln Asn  
 165 170 175  
 Ile Pro Leu Asp Gly His Glu Ala Asp Glu Val Pro Glu Gly Leu Phe  
 180 185 190  
 Ala Pro Asp Asn Phe Gln Ala Leu Leu Glu Cys Arg Ile Asn Ser Gly  
 195 200 205  
 Glu Glu Val Leu Arg Lys Arg Phe Glu Ala Thr Ala Val Asn Thr Leu  
 210 215 220  
 Phe Cys Ser Lys Thr Gln Gln Arg His Met Leu Glu Ile Cys Glu Ser  
 225 230 235 240  
 Cys Ile Arg Glu Glu Thr Leu Arg Glu Val Arg Asp Ser His Phe Phe  
 245 250 255  
 Ser Ile Ile Thr Asp Asp Val Val Asp Ile Ala Gly Glu Glu His Leu  
 260 265 270  
 Pro Val Leu Val Arg Phe Val Asp Asp Ala His Asn Leu Arg Glu Glu  
 275 280 285  
 Phe Val Gly Phe Leu Pro Tyr Glu Ala Asp Ala Glu Ile Leu Ala Val  
 290 295 300  
 Lys Phe His Thr Thr Ile Thr Glu Lys Trp Gly Leu Asn Met Glu Tyr  
 305 310 315 320  
 Cys Arg Gly Gln Ala Tyr Ile Val Ser Ser Gly Phe Ser Ser Lys Met  
 325 330 335  
 Lys Val Val Ala Ser Arg Leu Leu Glu Lys Tyr Pro Gln Ala Val Tyr  
 340 345 350  
 Thr Leu Cys Ser Ser Cys Ala Leu Asn Ala Trp Leu Ala Lys Ser Val  
 355 360 365  
 Pro Val Ile Gly Val Ser Val Ala Leu Gly Thr Ile Glu Glu Val Cys  
 370 375 380  
 Ser Phe Phe His Arg Ser Pro Gln Leu Leu Leu Glu Leu Asp Ser Val  
 385 390 395 400  
 Ile Ser Val Leu Phe Gln Asn Ser Glu Glu Arg Ala Lys Glu Leu Lys  
 405 410 415  
 Glu Ile Cys His Ser Gln Trp Thr Gly Arg His Asp Ala Phe Glu Ile  
 420 425 430  
 Leu Val Asp Leu Leu Gln Ala Leu Val Leu Cys Leu Asp Gly Ile Ile  
 435 440 445  
 Asn Ser Asp Thr Asn Val Arg Trp Asn Asn Tyr Ile Ala Gly Arg Ala  
 450 455 460  
 Phe Val Leu Cys Ser Ala Val Thr Asp Phe Asp Phe Ile Val Thr Ile  
 465 470 475 480  
 Val Val Leu Lys Asn Val Leu Ser Phe Thr Arg Ala Phe Gly Lys Asn  
 485 490 495  
 Leu Gln Gly Gln Thr Ser Asp Val Phe Phe Ala Ala Ser Ser Leu Thr  
 500 505 510

Ala Val Leu His Ser Leu Asn Glu Val Met Glu Asn Ile Glu Val Tyr  
 515 520 525  
 His Glu Phe Trp Phe Glu Glu Ala Thr Asn Leu Ala Thr Lys Leu Asp  
 530 535 540  
 Ile Gln Met Lys Leu Pro Gly Lys Phe Arg Arg Ala Gln Gln Gly Asn  
 545 550 555 560  
 Leu Glu Ser Gln Leu Thr Ser Glu Ser Tyr Tyr Lys Asp Thr Leu Ser  
 565 570 575  
 Val Pro Thr Val Glu His Ile Ile Gln Glu Leu Lys Asp Ile Phe Ser  
 580 585 590  
 Glu Gln His Leu Lys Ala Leu Lys Cys Leu Ser Leu Val Pro Ser Val  
 595 600 605  
 Met Gly Gln Leu Lys Phe Asn Thr Ser Glu Glu His His Ala Asp Met  
 610 615 620  
 Tyr Arg Ser Asp Leu Pro Asn Pro Asp Thr Leu Ser Ala Glu Leu His  
 625 630 635 640  
 Cys Trp Arg Ile Lys Trp Lys His Arg Gly Lys Asp  
 645 650

<210> 108  
 <211> 180  
 <212> PRT  
 <213> Rattus norvegicus

<220>  
 <223> RAT THAP

<221> UNSURE  
 <222> 95  
 <223> Xaa = any of the twenty amino acids

<400> 108  
 Arg Gln Cys Cys Asn Arg Tyr Ser Ser Arg Arg Lys Gln Leu Thr Phe  
 1 5 10 15  
 His Arg Phe Pro Phe Ser Arg Pro Glu Leu Leu Arg Glu Trp Val Leu  
 20 25 30  
 Asn Ile Gly Arg Ala Asp Phe Lys Pro Lys Gln His Thr Val Ile Cys  
 35 40 45  
 Ser Glu His Phe Arg Pro Glu Cys Phe Ser Ala Phe Gly Asn Arg Lys  
 50 55 60  
 Asn Leu Lys His Asn Ala Val Pro Thr Val Phe Ala Phe Gln Asn Pro  
 65 70 75 80  
 Ala Gln Val Cys Pro Glu Val Gly Ala Gly Gly Asp Ser Ser Xaa Arg  
 85 90 95  
 Asn Met Asp Ala Thr Leu Glu Glu Leu Gln Ser Pro Asn Thr Glu Gly  
 100 105 110  
 Pro Met Gln Gln Val Leu Pro Asp Arg Gln Ala Thr Glu Ala Met Glu  
 115 120 125  
 Ala Ala Gly Leu Pro Ala Gly Pro Leu Gly Leu Lys Arg Pro Leu Pro  
 130 135 140  
 Gly Gln Pro Ser Asp His Ser Tyr Ala Leu Leu Asp Leu Asp Thr Leu  
 145 150 155 160  
 Lys Lys Lys Leu Phe Leu Thr Leu Lys Glu Asn Lys Arg Leu Arg Lys  
 165 170 175  
 Arg Leu Lys Ala  
 180

<210> 109  
 <211> 82

&lt;212&gt; PRT

&lt;213&gt; Rattus norvegicus

&lt;400&gt; 109

Met Val Lys Cys Cys Ser Ala Ile Gly Cys Ala Ser Arg Cys Leu Pro  
 1 5 10 15  
 Asn Ser Lys Leu Lys Gly Leu Thr Phe His Val Phe Pro Thr Asp Glu  
 20 25 30  
 Asn Ile Lys Arg Lys Trp Val Leu Ala Met Lys Arg Leu Asp Val Asn  
 35 40 45  
 Thr Ala Gly Ile Trp Glu Pro Ser Leu Gln Pro Glu Ser Phe Tyr Phe  
 50 55 60  
 Ile Phe Met Glu Asn Leu Phe Phe Ile Leu Pro Pro Gln Leu Ser His  
 65 70 75 80  
 Ala Val

&lt;210&gt; 110

&lt;211&gt; 309

&lt;212&gt; PRT

&lt;213&gt; Rattus norvegicus

&lt;400&gt; 110

Met Pro Arg His Cys Ser Ala Ala Gly Cys Cys Thr Arg Asp Thr Arg  
 1 5 10 15  
 Glu Thr Arg Asn Arg Gly Ile Ser Phe His Arg Leu Pro Lys Lys Asp  
 20 25 30  
 Asn Pro Arg Arg Gly Leu Trp Leu Ala Asn Cys Gln Arg Leu Asp Pro  
 35 40 45  
 Ser Gly Gln Gly Leu Trp Asp Pro Thr Ser Glu Tyr Ile Tyr Phe Cys  
 50 55 60  
 Ser Lys His Phe Glu Glu Asn Cys Phe Glu Leu Val Gly Ile Ser Gly  
 65 70 75 80  
 Tyr His Arg Leu Lys Glu Gly Ala Val Pro Thr Ile Phe Glu Ser Phe  
 85 90 95  
 Ser Lys Leu Arg Arg Thr Ala Lys Thr Lys Val His Gly Tyr Pro Pro  
 100 105 110  
 Gly Leu Pro Asp Val Ser Arg Leu Arg Arg Cys Arg Lys Arg Cys Ser  
 115 120 125  
 Glu Arg Gln Gly Pro Thr Ile Pro Phe Ser Pro Pro Pro Arg Ala Asp  
 130 135 140  
 Ile Ile Arg Phe Pro Val Glu Glu Ala Ser Ala Pro Ala Thr Leu Pro  
 145 150 155 160  
 Ala Ser Pro Ala Ala Arg Leu Asp Pro Gly Leu Asn Ser Pro Phe Ser  
 165 170 175  
 Asp Leu Leu Gly Pro Leu Gly Ala Gln Ala Asp Glu Ala Gly Cys Ser  
 180 185 190  
 Ala Gln Pro Ser Pro Glu Gln His Pro Ser Pro Leu Glu Pro Gln His  
 195 200 205  
 Val Ser Pro Ser Thr Tyr Met Leu Arg Leu Pro Pro Pro Ala Gly Ala  
 210 215 220  
 Tyr Ile Gln Asn Glu His Ser Tyr Gln Val Gly Ser Ala Leu Leu Trp  
 225 230 235 240  
 Lys Arg Arg Ala Glu Ala Ala Leu Asp Ala Leu Asp Lys Thr Gln Arg  
 245 250 255  
 Gln Leu Gln Ala Cys Lys Arg Arg Glu Gln Arg Leu Arg Leu Arg Leu  
 260 265 270  
 Thr Lys Leu Gln Gln Glu Arg Ala Arg Glu Lys Arg Ala Gln Ala Asp  
 275 280 285  
 Ala Arg Gln Thr Leu Lys Asp His Val Gln Asp Phe Ala Met Gln Leu

290  
Ser Ser Ser Met Ala  
305

300

<210> 111  
<211> 142  
<212> PRT  
<213> Rattus norvegicus

<400> 111  
Met Pro Asn Phe Cys Ala Ala Pro Asn Cys Thr Arg Lys Ser Thr Gln  
1 5 10 15  
Ser Asp Leu Ala Phe Phe Arg Phe Pro Arg Asp Pro Ala Arg Cys Gln  
20 25 30  
Lys Trp Val Glu Asn Cys Arg Arg Ala Asp Leu Glu Asp Lys Thr Pro  
35 40 45  
Asp Gln Leu Asn Lys His Tyr Arg Leu Cys Ala Lys His Phe Glu Thr  
50 55 60  
Ser Met Ile Cys Arg Thr Ser Pro Tyr Arg Thr Val Leu Arg Asp Asn  
65 70 75 80  
Ala Ile Pro Thr Ile Phe Asp Leu Thr Ser His Leu Asn Asn Pro His  
85 90 95  
Ser Arg His Arg Lys Arg Ile Lys Glu Leu Ser Glu Asp Glu Ile Arg  
100 105 110  
Thr Leu Lys Gln Lys Lys Ile Glu Glu Thr Ser Glu Gln Glu Gln Gly  
115 120 125  
Thr Asn Ser Asn Ala Gln Tyr Pro Ser Ala Glu Val Gly Asn  
130 135 140

<210> 112  
<211> 104  
<212> PRT  
<213> Sus scrofa

<400> 112  
Met Val Lys Cys Cys Ser Ala Ile Gly Cys Ala Ser Arg Cys Leu Pro  
1 5 10 15  
Asn Ser Lys Leu Lys Gly Leu Thr Phe His Val Phe Pro Thr Asp Glu  
20 25 30  
Lys Val Lys Arg Lys Trp Val Leu Ala Met Lys Arg Leu Asp Val Asn  
35 40 45  
Ala Ala Gly Met Trp Glu Pro Lys Lys Gly Asp Val Leu Cys Ser Arg  
50 55 60  
His Phe Lys Lys Thr Asp Phe Asp Arg Thr Thr Pro Asn Ile Lys Leu  
65 70 75 80  
Lys Pro Gly Val Ile Pro Ser Ile Phe Asp Ser Pro Ser His Leu Thr  
85 90 95  
Gly Glu Glu Arg Lys Ala Pro Leu  
100

<210> 113  
<211> 235  
<212> PRT  
<213> Sus scrofa

<220>  
<221> UNSURE  
<222> 57, 124, 192

<223> Xaa = any of the twenty amino acids

<400> 113

```

Met Pro Arg His Cys Ser Ala Ala Gly Cys Cys Thr Arg Asp Thr Arg
1          5          10          15
Glu Thr Arg Asn Arg Gly Ile Ser Phe His Arg Leu Pro Lys Lys Asp
20          25          30
Asn Pro Arg Arg Gly Leu Trp Leu Ala Asn Cys Gln Arg Leu Asp Pro
35          40          45
Ser Gly Gln Gly Leu Trp Asp Pro Xaa Ser Glu Tyr Ile Tyr Phe Cys
50          55          60
Ser Lys His Phe Glu Glu Asn Cys Phe Glu Leu Val Gly Ile Ser Gly
65          70          75          80
Tyr His Arg Leu Lys Glu Gly Ala Val Pro Thr Ile Phe Glu Ser Phe
85          90          95
Ser Lys Leu Arg Arg Thr Ala Lys Thr Lys Gly His Ser Tyr Pro Pro
100          105          110
Gly Pro Pro Asp Val Ser Arg Leu Arg Arg Cys Xaa Lys Arg Cys Ser
115          120          125
Glu Gly Arg Gly Pro Thr Thr Pro Phe Ser Pro Pro Pro Ala Asp
130          135          140
Val Thr Cys Phe Pro Val Glu Glu Ala Ser Ala Pro Ala Ala Leu Ser
145          150          155          160
Ala Ser Pro Thr Gly Arg Leu Glu Pro Gly Leu Ser Ser Pro Phe Ser
165          170          175
Asp Leu Leu Gly Pro Leu Gly Ala Gln Ala Asp Glu Ala Gly Cys Xaa
180          185          190
Thr Gln Pro Ser Pro Glu Arg Glu Pro Glu Arg Gln Pro Ser Pro Leu
195          200          205
Glu Pro Arg Pro Val Ser Pro Ser Ala Tyr Met Leu Arg Leu Pro Pro
210          215          220
Pro Ala Gly Ala Tyr Ile Gln Asn Glu His Ser
225          230          235

```

<210> 114

<211> 149

<212> PRT

<213> Sus scrofa

<400> 114

```

Met Thr Arg Ser Cys Ser Ala Val Gly Cys Ser Thr Arg Asp Thr Val
1          5          10          15
Leu Ser Arg Glu Arg Gly Leu Ser Phe His Gln Phe Pro Thr Asp Thr
20          25          30
Ile Gln Arg Ser Gln Trp Ile Arg Ala Val Asn Arg Met Asp Pro Arg
35          40          45
Ser Lys Lys Ile Trp Ile Pro Gly Pro Gly Ala Met Leu Cys Ser Lys
50          55          60
His Phe Gln Glu Ser Asp Phe Glu Ser Tyr Gly Ile Arg Arg Lys Leu
65          70          75          80
Lys Lys Gly Ala Val Pro Ser Val Ser Leu Tyr Lys Val Leu Gln Gly
85          90          95
Ala His Leu Lys Gly Lys Ala Arg Gln Lys Ile Leu Lys Gln Pro Leu
100          105          110
Pro Asp Asn Ser Gln Glu Val Ala Thr Glu Asp His Asn Tyr Ser Leu
115          120          125
Lys Gly Pro Leu Thr Ile Gly Ala Glu Lys Leu Ala Glu Val Gln Gln
130          135          140
Met Leu Gln Val Ser
145

```



<210> 115  
 <211> 43  
 <212> PRT  
 <213> Mus musculus

<400> 115  
 Val Leu Glu Asp Val Ala Ala Ala Glu Gln Gly Leu Arg Glu Leu Gln  
 1 5 10 15  
 Arg Gly Arg Arg Gln Cys Arg Glu Arg Val Cys Ala Leu Arg Ala Ala  
 20 25 30  
 Ala Glu Gln Arg Glu Ala Arg Cys Arg Asp Gly  
 35 40

<210> 116  
 <211> 45  
 <212> PRT  
 <213> Mus musculus

<400> 116  
 Gln Leu Glu Gln Gln Val Glu Lys Leu Arg Lys Lys Leu Lys Thr Ala  
 1 5 10 15  
 Gln Gln Arg Cys Arg Arg Gln Glu Arg Gln Leu Glu Lys Leu Lys Glu  
 20 25 30  
 Val Val His Phe Gln Arg Glu Lys Asp Asp Ala Ser Glu  
 35 40 45

<210> 117  
 <211> 45  
 <212> PRT  
 <213> Homo sapiens

<400> 117  
 Gln Leu Glu Gln Gln Val Glu Lys Leu Arg Lys Lys Leu Lys Thr Ala  
 1 5 10 15  
 Gln Gln Arg Cys Arg Arg Gln Glu Arg Gln Leu Glu Lys Leu Lys Glu  
 20 25 30  
 Val Val His Phe Gln Lys Glu Lys Asp Asp Val Ser Glu  
 35 40 45

<210> 118  
 <211> 342  
 <212> PRT  
 <213> Homo sapiens

<400> 118  
 Met Ala Thr Gly Gly Tyr Arg Thr Ser Ser Gly Leu Gly Gly Ser Thr  
 1 5 10 15  
 Thr Asp Phe Leu Glu Glu Trp Lys Ala Lys Arg Glu Lys Met Arg Ala  
 20 25 30  
 Lys Gln Asn Pro Pro Gly Pro Ala Pro Pro Gly Gly Gly Ser Ser Asp  
 35 40 45  
 Ala Ala Gly Lys Pro Pro Ala Gly Ala Leu Gly Thr Pro Ala Ala Ala  
 50 55 60  
 Ala Ala Asn Glu Leu Asn Asn Asn Leu Pro Gly Gly Ala Pro Ala Ala  
 65 70 75 80  
 Pro Ala Val Pro Gly Pro Gly Gly Val Asn Cys Ala Val Gly Ser Ala

85 90 95  
 Met Leu Thr Arg Ala Pro Pro Ala Arg Gly Pro Arg Arg Ser Glu Asp  
 100 105 110  
 Glu Pro Pro Ala Ala Ser Ala Ser Ala Ala Pro Pro Pro Gln Arg Asp  
 115 120 125  
 Glu Glu Glu Pro Asp Gly Val Pro Glu Lys Gly Lys Ser Ser Gly Pro  
 130 135 140  
 Ser Ala Arg Lys Gly Lys Gly Gln Ile Glu Lys Arg Lys Leu Arg Glu  
 145 150 155 160  
 Lys Arg Arg Ser Thr Gly Val Val Asn Ile Pro Ala Ala Glu Cys Leu  
 165 170 175  
 Asp Glu Tyr Glu Asp Asp Glu Ala Gly Gln Lys Glu Arg Lys Arg Glu  
 180 185 190  
 Asp Ala Ile Thr Gln Gln Asn Thr Ile Gln Asn Glu Ala Val Asn Leu  
 195 200 205  
 Leu Asp Pro Gly Ser Ser Tyr Leu Leu Gln Glu Pro Pro Arg Thr Val  
 210 215 220  
 Ser Gly Arg Tyr Lys Ser Thr Thr Ser Val Ser Glu Glu Asp Val Ser  
 225 230 235 240  
 Ser Arg Tyr Ser Arg Thr Asp Arg Ser Gly Phe Pro Arg Tyr Asn Arg  
 245 250 255  
 Asp Ala Asn Val Ser Gly Thr Leu Val Ser Ser Ser Thr Leu Glu Lys  
 260 265 270  
 Lys Ile Glu Asp Leu Glu Lys Glu Val Val Thr Glu Arg Gln Glu Asn  
 275 280 285  
 Leu Arg Leu Val Arg Leu Met Gln Asp Lys Glu Glu Met Ile Gly Lys  
 290 295 300  
 Leu Lys Glu Glu Ile Asp Leu Leu Asn Arg Asp Leu Asp Asp Ile Glu  
 305 310 315 320  
 Asp Glu Asn Glu Gln Leu Lys Gln Glu Asn Lys Thr Leu Leu Lys Val  
 325 330 335  
 Val Gly Gln Leu Thr Arg  
 340

<210> 119  
 <211> 134  
 <212> PRT  
 <213> Homo sapiens

<400> 119  
 Met Ala Gln Ser Leu Ala Leu Ser Leu Leu Ile Leu Val Leu Ala Phe  
 1 5 10 15  
 Gly Ile Pro Arg Thr Gln Gly Ser Asp Gly Gly Ala Gln Asp Cys Cys  
 20 25 30  
 Leu Lys Tyr Ser Gln Arg Lys Ile Pro Ala Lys Val Val Arg Ser Tyr  
 35 40 45  
 Arg Lys Gln Glu Pro Ser Leu Gly Cys Ser Ile Pro Ala Ile Leu Phe  
 50 55 60  
 Leu Pro Arg Lys Arg Ser Gln Ala Glu Leu Cys Ala Asp Pro Lys Glu  
 65 70 75 80  
 Leu Trp Val Gln Gln Leu Met Gln His Leu Asp Lys Thr Pro Ser Pro  
 85 90 95  
 Gln Lys Pro Ala Gln Gly Cys Arg Lys Asp Arg Gly Ala Ser Lys Thr  
 100 105 110  
 Gly Lys Lys Gly Lys Gly Ser Lys Gly Cys Lys Arg Thr Glu Arg Ser  
 115 120 125  
 Gln Thr Pro Lys Gly Pro  
 130

&lt;210&gt; 120

&lt;211&gt; 766

&lt;212&gt; PRT

<213> *Drosophila melanogaster*

&lt;400&gt; 120

```

Met Lys Tyr Cys Lys Phe Cys Cys Lys Ala Val Thr Gly Val Lys Leu
 1      5      10      15
Ile His Val Pro Lys Cys Ala Ile Lys Arg Lys Leu Trp Glu Gln Ser
      20      25      30
Leu Gly Cys Ser Leu Gly Glu Asn Ser Gln Ile Cys Asp Thr His Phe
      35      40      45
Asn Asp Ser Gln Trp Lys Ala Ala Pro Ala Lys Gly Gln Thr Phe Lys
      50      55      60
Arg Arg Arg Leu Asn Ala Asp Ala Val Pro Ser Lys Val Ile Glu Pro
      65      70      75      80
Glu Pro Glu Lys Ile Lys Glu Gly Tyr Thr Ser Gly Ser Thr Gln Thr
      85      90      95
Glu Ser Cys Ser Leu Phe Asn Glu Asn Lys Ser Leu Arg Glu Lys Ile
      100      105      110
Arg Thr Leu Glu Tyr Glu Met Arg Arg Leu Glu Gln Gln Leu Arg Glu
      115      120      125
Ser Gln Gln Leu Glu Glu Ser Leu Arg Lys Ile Phe Thr Asp Thr Gln
      130      135      140
Ile Arg Ile Leu Lys Asn Gly Gly Gln Arg Ala Thr Phe Asn Ser Asp
      145      150      155      160
Asp Ile Ser Thr Ala Ile Cys Leu His Thr Ala Gly Pro Arg Ala Tyr
      165      170      175
Asn His Leu Tyr Lys Lys Gly Phe Pro Leu Pro Ser Arg Thr Thr Leu
      180      185      190
Tyr Arg Trp Leu Ser Asp Val Asp Ile Lys Arg Gly Cys Leu Asp Val
      195      200      205
Val Ile Asp Leu Met Asp Ser Asp Gly Val Asp Asp Ala Asp Lys Leu
      210      215      220
Cys Val Leu Ala Phe Asp Glu Met Lys Val Ala Ala Ala Phe Glu Tyr
      225      230      235      240
Asp Ser Ser Ala Asp Ile Val Tyr Glu Pro Ser Asp Tyr Val Gln Leu
      245      250      255
Ala Ile Val Arg Gly Leu Lys Lys Ser Trp Lys Gln Pro Val Phe Phe
      260      265      270
Asp Phe Asn Thr Arg Met Asp Pro Asp Thr Leu Asn Asn Ile Leu Arg
      275      280      285
Lys Leu His Arg Lys Gly Tyr Leu Val Val Ala Ile Val Ser Asp Leu
      290      295      300
Gly Thr Gly Asn Gln Lys Leu Trp Thr Glu Leu Gly Ile Ser Glu Ser
      305      310      315      320
Lys Thr Trp Phe Ser His Pro Ala Asp Asp His Leu Lys Ile Phe Val
      325      330      335
Phe Ser Asp Thr Pro His Leu Ile Lys Leu Val Arg Asn His Tyr Val
      340      345      350
Asp Ser Gly Leu Thr Ile Asn Gly Lys Lys Leu Thr Lys Lys Thr Ile
      355      360      365
Gln Glu Ala Leu His Leu Cys Asn Lys Ser Asp Leu Ser Ile Leu Phe
      370      375      380
Lys Ile Asn Glu Asn His Ile Asn Val Arg Ser Leu Ala Lys Gln Lys
      385      390      395      400
Val Lys Leu Ala Thr Gln Leu Phe Ser Asn Thr Thr Ala Ser Ser Ile
      405      410      415
Arg Arg Cys Tyr Ser Leu Gly Tyr Asp Ile Glu Asn Ala Thr Glu Thr
      420      425      430
Ala Asp Phe Phe Lys Leu Met Asn Asp Trp Phe Asp Ile Phe Asn Ser

```

435 440 445  
 Lys Leu Ser Thr Ser Asn Cys Ile Glu Cys Ser Gln Pro Tyr Gly Lys  
 450 455 460  
 Gln Leu Asp Ile Gln Asn Asp Ile Leu Asn Arg Met Ser Glu Ile Met  
 465 470 475 480  
 Arg Thr Gly Ile Leu Asp Lys Pro Lys Arg Leu Pro Phe Gln Lys Gly  
 485 490 495  
 Ile Ile Val Asn Asn Ala Ser Leu Asp Gly Leu Tyr Lys Tyr Leu Gln  
 500 505 510  
 Glu Asn Phe Ser Met Gln Tyr Ile Leu Thr Ser Arg Leu Asn Gln Asp  
 515 520 525  
 Ile Val Glu His Phe Phe Gly Ser Met Arg Ser Arg Gly Gly Gln Phe  
 530 535 540  
 Asp His Pro Thr Pro Leu Gln Phe Lys Tyr Arg Leu Arg Lys Tyr Ile  
 545 550 555 560  
 Ile Ala Arg Asn Thr Glu Met Leu Arg Asn Ser Gly Asn Ile Glu Glu  
 565 570 575  
 Gly Met Thr Asn Leu Lys Glu Cys Val Asn Lys Asn Val Ile Pro Asp  
 580 585 590  
 Asn Ser Glu Ser Trp Leu Asn Leu Asp Phe Ser Ser Lys Glu Asn Glu  
 595 600 605  
 Asn Lys Ser Lys Asp Asp Glu Pro Val Asp Asp Glu Pro Val Asp Glu  
 610 615 620  
 Met Leu Ser Asn Ile Asp Phe Thr Glu Met Asp Glu Leu Thr Glu Asp  
 625 630 635 640  
 Ala Met Glu Tyr Ile Ala Gly Tyr Val Ile Lys Lys Leu Arg Ile Ser  
 645 650 655  
 Asp Lys Val Lys Glu Asn Leu Thr Phe Thr Tyr Val Asp Glu Val Ser  
 660 665 670  
 His Gly Gly Leu Ile Lys Pro Ser Glu Lys Phe Gln Glu Lys Leu Lys  
 675 680 685  
 Glu Leu Glu Cys Ile Phe Leu His Tyr Thr Asn Asn Asn Asn Phe Glu  
 690 695 700  
 Ile Thr Asn Asn Val Lys Glu Lys Leu Ile Leu Ala Ala Arg Asn Val  
 705 710 715 720  
 Asp Val Asp Lys Gln Val Lys Ser Phe Tyr Phe Lys Ile Arg Ile Tyr  
 725 730 735  
 Phe Arg Ile Lys Tyr Phe Asn Lys Lys Ile Glu Ile Lys Asn Gln Lys  
 740 745 750  
 Gln Lys Leu Ile Gly Asn Ser Lys Leu Leu Lys Ile Lys Leu  
 755 760 765

<210> 121  
 <211> 103  
 <212> PRT  
 <213> Homo sapiens

<400> 121  
 Asp Glu Leu Cys Val Val Cys Gly Asp Lys Ala Thr Gly Tyr His Tyr  
 1 5 10 15  
 Arg Cys Ile Thr Cys Glu Gly Cys Lys Gly Phe Phe Arg Arg Thr Ile  
 20 25 30  
 Gln Lys Asn Leu His Pro Ser Tyr Ser Cys Lys Tyr Glu Gly Lys Cys  
 35 40 45  
 Val Ile Asp Lys Val Thr Arg Asn Gln Cys Gln Glu Cys Arg Phe Lys  
 50 55 60  
 Lys Cys Ile Tyr Val Gly Met Ala Thr Asp Leu Val Leu Asp Asp Ser  
 65 70 75 80  
 Lys Arg Leu Ala Lys Arg Lys Leu Ile Glu Glu Asn Arg Glu Lys Arg  
 85 90 95

Arg Arg Glu Glu Leu Glu Lys  
100

<210> 122

<211> 81

<212> PRT

<213> Homo sapiens

<400> 122

Met	Lys	Pro	Ala	Arg	Pro	Cys	Leu	Val	Cys	Ser	Asp	Glu	Ala	Ser	Gly
1				5					10					15	
Cys	His	Tyr	Gly	Val	Leu	Thr	Cys	Gly	Ser	Cys	Lys	Val	Phe	Phe	Lys
			20					25					30		
Arg	Ala	Val	Glu	Gly	Gln	His	Asn	Tyr	Leu	Cys	Ala	Gly	Arg	Asn	Asp
		35					40					45			
Cys	Ile	Ile	Asp	Lys	Ile	Arg	Arg	Lys	Asn	Cys	Pro	Ala	Cys	Arg	Tyr
	50					55				60					
Arg	Lys	Cys	Leu	Gln	Ala	Gly	Met	Asn	Leu	Glu	Ala	Arg	Lys	Thr	Lys
65					70					75					80
Lys															

<210> 123

<211> 89

<212> PRT

<213> Homo sapiens

<400> 123

Met	Val	Gln	Ser	Cys	Ser	Ala	Tyr	Gly	Cys	Lys	Asn	Arg	Tyr	Asp	Lys
1				5					10					15	
Asp	Lys	Pro	Val	Ser	Phe	His	Lys	Phe	Pro	Leu	Thr	Arg	Pro	Ser	Leu
			20					25					30		
Cys	Lys	Glu	Trp	Glu	Ala	Ala	Val	Arg	Arg	Lys	Asn	Phe	Lys	Pro	Thr
		35					40				45				
Lys	Tyr	Ser	Ser	Ile	Cys	Ser	Glu	His	Phe	Thr	Pro	Asp	Cys	Phe	Lys
	50					55				60					
Arg	Glu	Cys	Asn	Asn	Lys	Leu	Leu	Lys	Glu	Asn	Ala	Val	Pro	Thr	Ile
65					70					75					80
Phe	Leu	Cys	Thr	Glu	Pro	His	Asp	Lys							
					85										

<210> 124

<211> 85

<212> PRT

<213> Drosophila melanogaster

<400> 124

Met	Lys	Tyr	Cys	Lys	Phe	Cys	Cys	Lys	Ala	Val	Thr	Gly	Val	Lys	Leu
1				5					10					15	
Ile	His	Val	Pro	Lys	Cys	Ala	Ile	Lys	Arg	Lys	Leu	Trp	Glu	Gln	Ser
			20					25					30		
Leu	Gly	Cys	Ser	Leu	Gly	Glu	Asn	Ser	Gln	Ile	Cys	Asp	Thr	His	Phe
		35					40				45				
Asn	Asp	Ser	Gln	Trp	Lys	Ala	Ala	Pro	Ala	Lys	Gly	Gln	Thr	Phe	Lys
	50					55				60					
Arg	Arg	Arg	Leu	Asn	Ala	Asp	Ala	Val	Pro	Ser	Lys	Val	Ile	Glu	Pro
65					70					75					80
Glu	Pro	Glu	Lys	Ile											

85

<210> 125  
 <211> 58  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> THAP Domain consensus

<221> UNSURE  
 <222> 2-3, 7, 9, 13-17, 19, 21-23, 25-26, 28, 35, 38-39, 41, 45-50,  
 52, 55-56  
 <223> Xaa = any of the twenty amino acids

<400> 125  
 Met Val Xaa Xaa Cys Ser Xaa Tyr Xaa Cys Lys Asn Xaa Xaa Xaa Xaa  
 1 5 10 15  
 Xaa Lys Xaa Val Xaa Xaa Xaa Lys Xaa Xaa Leu Xaa Arg Pro Ser Leu  
 20 25 30  
 Cys Lys Xaa Trp Glu Xaa Xaa Val Xaa Arg Lys Asn Xaa Xaa Xaa Xaa  
 35 40 45  
 Xaa Xaa Ser Xaa Ile Cys Xaa Xaa His Phe  
 50 55

<210> 126  
 <211> 89  
 <212> PRT  
 <213> Homo sapiens

<400> 126  
 Met Val Gln Ser Cys Ser Ala Tyr Gly Cys Lys Asn Arg Tyr Asp Lys  
 1 5 10 15  
 Asp Lys Pro Val Ser Phe His Lys Phe Pro Leu Thr Arg Pro Ser Leu  
 20 25 30  
 Cys Lys Glu Trp Glu Ala Ala Val Arg Arg Lys Asn Phe Lys Pro Thr  
 35 40 45  
 Lys Tyr Ser Ser Ile Cys Ser Glu His Phe Thr Pro Asp Cys Phe Lys  
 50 55 60  
 Arg Glu Cys Asn Asn Lys Leu Leu Lys Glu Asn Ala Val Pro Thr Ile  
 65 70 75 80  
 Phe Leu Cys Thr Glu Pro His Asp Lys  
 85

<210> 127  
 <211> 89  
 <212> PRT  
 <213> Homo sapiens

<400> 127  
 Met Pro Lys Ser Cys Ala Ala Arg Gln Cys Cys Asn Arg Tyr Ser Ser  
 1 5 10 15  
 Arg Arg Lys Gln Leu Thr Phe His Arg Phe Pro Phe Ser Arg Pro Glu  
 20 25 30  
 Leu Leu Lys Glu Trp Val Leu Asn Ile Gly Arg Gly Asn Phe Lys Pro  
 35 40 45  
 Lys Gln His Thr Val Ile Cys Ser Glu His Phe Arg Pro Glu Cys Phe  
 50 55 60

Ser Ala Phe Gly Asn Arg Lys Asn Leu Lys His Asn Ala Val Pro Thr  
 65 70 75 80  
 Val Phe Ala Phe Gln Asp Pro Thr Gln  
 85

<210> 128  
 <211> 90  
 <212> PRT  
 <213> Homo sapiens

<400> 128  
 Met Pro Arg Tyr Cys Ala Ala Ile Cys Cys Lys Asn Arg Arg Gly Arg  
 1 5 10 15  
 Asn Asn Lys Asp Arg Lys Leu Ser Phe Tyr Pro Phe Pro Leu His Asp  
 20 25 30  
 Lys Glu Arg Leu Glu Lys Trp Leu Lys Asn Met Lys Arg Asp Ser Trp  
 35 40 45  
 Val Pro Ser Lys Tyr Gln Phe Leu Cys Ser Asp His Phe Thr Pro Asp  
 50 55 60  
 Ser Leu Asp Ile Arg Trp Gly Ile Arg Tyr Leu Lys Gln Thr Ala Val  
 65 70 75 80  
 Pro Thr Ile Phe Ser Leu Pro Glu Asp Asn  
 85 90

<210> 129  
 <211> 92  
 <212> PRT  
 <213> Homo sapiens

<400> 129  
 Met Pro Lys Tyr Cys Arg Ala Pro Asn Cys Ser Asn Thr Ala Gly Arg  
 1 5 10 15  
 Leu Gly Ala Asp Asn Arg Pro Val Ser Phe Tyr Lys Phe Pro Leu Lys  
 20 25 30  
 Asp Gly Pro Arg Leu Gln Ala Trp Leu Gln His Met Gly Cys Glu His  
 35 40 45  
 Trp Val Pro Ser Cys His Gln His Leu Cys Ser Glu His Phe Thr Pro  
 50 55 60  
 Ser Cys Phe Gln Trp Arg Trp Gly Val Arg Tyr Leu Arg Pro Asp Ala  
 65 70 75 80  
 Val Pro Ser Ile Phe Ser Arg Gly Pro Pro Ala Lys  
 85 90

<210> 130  
 <211> 90  
 <212> PRT  
 <213> Homo sapiens

<400> 130  
 Met Val Ile Cys Cys Ala Ala Val Asn Cys Ser Asn Arg Gln Gly Lys  
 1 5 10 15  
 Gly Glu Lys Arg Ala Val Ser Phe His Arg Phe Pro Leu Lys Asp Ser  
 20 25 30  
 Lys Arg Leu Ile Gln Trp Leu Lys Ala Val Gln Arg Asp Asn Trp Thr  
 35 40 45  
 Pro Thr Lys Tyr Ser Phe Leu Cys Ser Glu His Phe Thr Lys Asp Ser  
 50 55 60  
 Phe Ser Lys Arg Leu Glu Asp Gln His Arg Leu Leu Lys Pro Thr Ala

```
<210> 131
<211> 89
<212> PRT
<213> Homo sapiens
```

```
<210> 132
<211> 90
<212> PRT
<213> Homo sapiens
```

```
<210> 133
<211> 97
<212> PRT
<213> Homo sapiens
```

-59-



Tyr His Arg Leu Lys Glu Gly Ala Val Pro Thr Ile Phe Glu Ser Phe  
                     85                    90                    95  
 Ser

<210> 134  
 <211> 92  
 <212> PRT  
 <213> Homo sapiens

<400> 134  
 Met Thr Arg Ser Cys Ser Ala Val Gly Cys Ser Thr Arg Asp Thr Val  
   1                    5                    10                    15  
 Leu Ser Arg Glu Arg Gly Leu Ser Phe His Gln Phe Pro Thr Asp Thr  
                     20                    25                    30  
 Ile Gln Arg Ser Lys Trp Ile Arg Ala Val Asn Arg Val Asp Pro Arg  
                     35                    40                    45  
 Ser Lys Lys Ile Trp Ile Pro Gly Pro Gly Ala Ile Leu Cys Ser Lys  
                     50                    55                    60  
 His Phe Gln Glu Ser Asp Phe Glu Ser Tyr Gly Ile Arg Arg Lys Leu  
  65                    70                    75                    80  
 Lys Lys Gly Ala Val Pro Ser Val Ser Leu Tyr Lys  
                     85                    90

<210> 135  
 <211> 96  
 <212> PRT  
 <213> Homo sapiens

<400> 135  
 Met Val Lys Cys Cys Ser Ala Ile Gly Cys Ala Ser Arg Cys Leu Pro  
   1                    5                    10                    15  
 Asn Ser Lys Leu Lys Gly Leu Thr Phe His Val Phe Pro Thr Asp Glu  
                     20                    25                    30  
 Asn Ile Lys Arg Lys Trp Val Leu Ala Met Lys Arg Leu Asp Val Asn  
                     35                    40                    45  
 Ala Ala Gly Ile Trp Glu Pro Lys Lys Gly Asp Val Leu Cys Ser Arg  
                     50                    55                    60  
 His Phe Lys Lys Thr Asp Phe Asp Arg Ser Ala Pro Asn Ile Lys Leu  
  65                    70                    75                    80  
 Lys Pro Gly Val Ile Pro Ser Ile Phe Asp Ser Pro Tyr His Leu Gln  
                     85                    90                    95

<210> 136  
 <211> 90  
 <212> PRT  
 <213> Homo sapiens

<400> 136  
 Met Pro Gly Phe Thr Cys Cys Val Pro Gly Cys Tyr Asn Asn Ser His  
   1                    5                    10                    15  
 Arg Asp Lys Ala Leu His Phe Tyr Thr Phe Pro Lys Asp Ala Glu Leu  
                     20                    25                    30  
 Arg Arg Leu Trp Leu Lys Asn Val Ser Arg Ala Gly Val Ser Gly Cys  
                     35                    40                    45  
 Phe Ser Thr Phe Gln Pro Thr Thr Gly His Arg Leu Cys Ser Val His  
                     50                    55                    60  
 Phe Gln Gly Gly Arg Lys Thr Tyr Thr Val Arg Val Pro Thr Ile Phe

```
<210> 137
<211> 90
<212> PRT
<213> Homo sapiens
```

```
<210> 138
<211> 85
<212> PRT
<213> Drosophila melanogaster
```

```
<210> 139
<211> 63
<212> PRT
<213> Artificial Sequence
```

```
<221> UNSURE
<222> 4-5, 7, 9-10, 12, 15-20, 22, 24, 32, 35, 38-39, 42-43, 46-47,
49-51, 53-61, 63
<223> Xaa = any of the twenty amino acids
```

-61-

Xaa Xaa Xaa Xaa Lys Xaa Lys Xaa Val Ser Phe His Lys Phe Pro Xaa  
                   20                  25                  30  
 His Asp Xaa His Asp Xaa Xaa Arg Arg Xaa Xaa Trp Val Xaa Xaa Val  
                   35                  40                  45  
 Xaa Xaa Xaa Arg Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Trp Xaa  
                   50                  55                  60

&lt;210&gt; 140

&lt;211&gt; 17

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; DR-5-related sequence

&lt;400&gt; 140

gggcatacta ctggcaa

17

&lt;210&gt; 141

&lt;211&gt; 17

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; DR-5-related sequence

&lt;400&gt; 141

gggcaaactg tgggcat

17

&lt;210&gt; 142

&lt;211&gt; 17

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; DR-5-related sequence

&lt;400&gt; 142

gggcatacta ctggcaa

17

&lt;210&gt; 143

&lt;211&gt; 17

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; DR-5-related sequence

&lt;400&gt; 143

gggcaaacta ctggcaa

17

&lt;210&gt; 144

&lt;211&gt; 17

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; DR-5-related sequence

&lt;400&gt; 144

gggccagttc gttgcaa  
17  
<210> 145  
<211> 16  
<212> DNA  
<213> Artificial Sequence  
<220>  
<223> DR-5-related sequence  
<400> 145  
gggcatgtac tggcaa  
16  
<210> 146  
<211> 16  
<212> DNA  
<213> Artificial Sequence  
<220>  
<223> DR-5-related sequence  
<400> 146  
gggcaactgt gggcaa  
16  
<210> 147  
<211> 18  
<212> DNA  
<213> Artificial Sequence  
<220>  
<223> DR-5-related sequence  
<400> 147  
gggcaacact actggcaa  
18  
<210> 148  
<211> 17  
<212> DNA  
<213> Artificial Sequence  
<220>  
<223> DR-5-related sequence  
<400> 148  
gggcaaagta ctggcaa  
17  
<210> 149  
<211> 17  
<212> DNA  
<213> Artificial Sequence  
<220>  
<223> DR-5 consensus sequence  
<221> unsure  
<222> 7-11  
<223> n = any of the four nucleotides  
<400> 149  
gggcaannnn ntggcaa  
17

<210> 150  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> ER-11-related sequence

<400> 150  
ttgccagtac taagtgtggg caa 23

<210> 151  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> ER-11-related sequence

<400> 151  
ctgccagtac atagtgtggg caa 23

<210> 152  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> ER-11-related sequence

<400> 152  
ttgccagtac taagtgtggg caa 23

<210> 153  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> ER-11-related sequence

<400> 153  
ctgccagtag atactgtggg caa 23

<210> 154  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> ER-11-related sequence

<400> 154  
ttgccagtag ttaggtgtgg gcga 24

<210> 155  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>

<223> ER-11-related sequence  
 <400> 155  
 ttgccagtag ttagtgtggg caa 23  
 <210> 156  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> ER-11-related sequence  
 <400> 156  
 ttgccagtac ctactaaggg caa 23  
 <210> 157  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> ER-11-related sequence  
 <400> 157  
 ttgccagtag ttagtgtggg cag 23  
 <210> 158  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> ER-11-related sequence  
 <400> 158  
 ctgccagtag taagtgtggg cag 23  
 <210> 159  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> ER-11 consensus sequence  
 <221> unsure  
 <222> 7-17  
 <223> n = any of the four nucleotides  
 <400> 159  
 ttgccannnn nnnnnnnnggg caa 23  
 <210> 160  
 <211> 642  
 <212> DNA  
 <213> Homo sapiens  
 <400> 160  
 atggtgcagt cctgctccgc ctacggctgc aagaaccgct acgacaagga caagcccgtt 60  
 tctttccaca agtttccctct tactcgacct agtctttgta aagaatggga ggcagctgtc 120

```

agaagaaaaa actttaaaacc caccaagtat agcagtattt gttcagagca ctttactcca 180
gactgcttta agagagagtg caacaacaag ttactgaaag agaattgctgt gcccacaata 240
tttcttttgta ctgagccaca tgacaagaaa gaagatcttc tggagccaca ggaacagctt 300
ccccacctc ctttaccgcc tctgtttcc caggttgatg ctgctattgg attactaatg 360
ccgcctcttc agaccctgt taatctctca gttttctgtg accacaacta tactgtggag 420
gatacaatgc accagcggaa aaggattcat cagctagaac agcaagttga aaaactcaga 480
aagaagctca agaccgcaca gcagcgatgc agaaggcaag aacggcagct tgaaaaatta 540
aaggagggttg ttactttcca gaaagagaaa gacgacgtat cagaaagagg ttatgtgatt 600
ctaccaaagt actactttga aatagttgaa gtaccagcat aa 642

```

&lt;210&gt; 161

&lt;211&gt; 687

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 161

```

atgccgacca attgcgctgc ggcggtgtgt gccactacct acaacaagca cattaacatc 60
agcttccaca ggtttccttt ggatcctaaa agaagaaaag aatgggttcg cctgggttagg 120
cgcaaaaatt ttgtgccagg aaaacacact ttctttgtt caaagcactt tgaagcctcc 180
tgttttgacc taacaggaca aactcgacga cttaaaatgg atgctgttcc aaccattttt 240
gatttttgta cccatataaa gtctatgaaa ctcaagtcaa ggaatctttt gaagaaaaac 300
aacagttgtt ctccagctgg accatctaatt ttaaaatcaa acattagtag tcagcaagta 360
ctacttgaac acagctatgc ctttaggaat cctatggagg caaaaaagag gatcattaaa 420
ctggaaaaag aaatagcaag cttaagaaga aaaatgaaaa cttgcctaca aaaggaacgc 480
agagcaactc gaagatggat caaagccacg tgtttggtta agaatttaga agcaaatagt 540
gtattaccta aaggtacatc agaacacatg ttaccaactg ccttaagcag tcttcctttg 600
gaagatttta agatccttga acaagatcaa caagataaaa cactgctaag tctaaatcta 660
aaacagacca agagtacctt catttaa 687

```

&lt;210&gt; 162

&lt;211&gt; 720

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 162

```

atgccgaagt cgtgcgcggc ccggcagtg tgcaaccgct acagcagccg caggaagcag 60
ctcaccttcc accggtttcc gttcagccgc ccggagctgc tgaaggaatg ggtgctgaac 120
atcggccggg gcaacttcaa gcccagcag cacacggtca tctgctccga gcaactcccg 180
ccagagtgtc tcagcgctt tggaaccgc aagaacctaa agcacaatgc cgtgccacg 240
gtgttcgcct tcaggaccc cacacagcag gtgagggaga acacagaccc tgccagtga 300
agaggaaatg ccagctcttc tcagaaagaa aaggtcctcc ctgaggcggg ggccggagag 360
gacagtccgt ggagaaacat ggacactgca cttgaagagc ttcagtggcc cccaaatgcc 420
gaaggccacg taaaacaggt ctgcgccacg aggccgcaag caacagaggc tgttgcccg 480
ccgactggcc ctgcaggcct gagaaggacc cccaacaagc agccatctga tcacagctat 540
gcccttttg acttagattc cctgaagaaa aaactcttcc tcaacttgaa ggaaaatgaa 600
aagctccgga agcgcttgca ggccagagg ctggtgatgc gaaggatgtc cagccgcctc 660
cgtgcttgca aagggcacca gggactccag gccagacttg ggccagagca gcagagctga 720

```

&lt;210&gt; 163

&lt;211&gt; 1734

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 163

```

atggtgatct gctgtgcggc cgtgaactgc tccaaccggc agggaaaagg cgagaagcgc 60
gccgtctcct tccacaggtt cccctaaaag gactcaaac gtctaatacca atggttaaaa 120
gctgttcaga gggataactg gactccact aagtattcat ttctcttag tgagcatttc 180
accaaagaca gcttctccaa gaggtggag gaccagcatc gcctgctgaa gccacggcc 240
gtgccatcca tttccacct gaccgagaag aagagggggg ctggaggcca tggccgacc 300
cggagaaaaa atgccagcaa ggccacaggg ggtgtgagg gacactcgag tgccgccacc 360

```

```

ggcagaggag ctgcagggtg gtcaccgtcc tcgagtggaa acccgatggc caagccagag 420
tcccgcaggt tgaagcaagc tgctctgcaa ggtgaagcca caccagggc ggcccaggag 480
gccgccagcc aggagcaggc ccagcaagct ctggaacgga ctccaggaga tggactggcc 540
accatggtgg caggcagtca gggaaaagca gaagcgtctg ccacagatgc tggcgatgag 600
agcgccactt cctccatcga agggggcggt acagataaga gtggcatttc tatggatgac 660
tttacgcccc caggatctgg ggcgtgcaaa tttatcggct cacttcattc gtacagtttc 720
tcctctaagc acacccgaga aaggccatct gtcccccag agccattga ccgcaagagg 780
ctgaagaaag atgtggaacc aagctgcagt gggagcagcc tgggaccgga caagggcctg 840
gccagagacc ctcccagctc atcacttacc gcgacaccgc agaagccttc ccagagcccc 900
tctgcccctc ctgcccagct caccctaaag ccagccacgg aagccgtgca gaggcagcac 960
agcgacgcca gccccatgtc catcaacgag gtcactctgt cggcgtcagg ggcctgcaag 1020
ctcatcgact cactgcactc ctactgcttc tctcccggc agaacaagag ccagggtgtg 1080
tgctgcggg agcagggtgga gaagaagaag ggcagctga agagcctgcg gcagagggtc 1140
agccgctccg acagccaggt gcggaagcta caggagaagc tggatgagct gaggagagt 1200
agcgtccctt atccaagtag cctgctgtcg ccagcccgcg agcccccaa gatgaacca 1260
gtgggtggag cactgtcctg gatgctgggc acctggctgt cggaccacc tggagccggg 1320
acctaccccc cactgcagcc cttccagtag ctggaggagg ttcacatctc ccacgtgggc 1380
cagcccatgc tgaacttctc gttcaactcc ttccaccgg acacgcgcaa gccgatgcac 1440
agagagtgtg gcttcattcg cctcaagccc gacaccaaca aggtggcctt tgtcagcgcc 1500
cagaacacag gcgtggtgga agtggaggag ggcgaggtag acgggcagga gctgtgcac 1560
gcatccact ccacgcccag gatctccttc gccaggagc cccacgtaga gcagatcac 1620
cggaagtcca ggctgaattc tgaaggcaaa cttgagcaga cggtctccat ggcaaccacg 1680
acacagccaa tgactcagca tcttcacgtc acctacaaga aggtgacccc gtaa 1734

```

&lt;210&gt; 164

&lt;211&gt; 1188

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 164

```

atgccccgct attgcgcagc gatttgttgt aagaaccgcc ggggacgaaa caataaagac 60
cggaagctga gtttttatcc atttctctta catgacaaaag aaagactgga aaagtggta 120
aagaatatga agcgagattc atgggttccc agtaaatacc agtttctatg tagtgaccat 180
tttactcctg actctcttga catcagatgg ggtattcgat atttaaaaca aactgcagtt 240
ccaacaatat tttctttgcc tgaagacaat cagggaaaag acccttctaa aaaaaaatcc 300
cagaagaaaa acttgaaga tgagaaagaa gtatgcccaa aagccaagtc agaagaatca 360
tttgtattaa atgagacaaa gaaaaatata gttaacacag atgtgcccc tcaacatcca 420
gaattacttc attcatcttc cttggtaaag ccaccagctc caaaaacagg aagtatacaa 480
aataacatgt taactcttaa tctagttaaa caacatactg ggaaaccaga atctaccttg 540
gaaacatcag ttaaccaaga tacaggtaga ggtggttttc acacatgttt tgagaaatca 600
aattctacaa ctattacttt gacaacttca aattcagaaa gtattcatca atctttggaa 660
actcaagaag ttcttgaagt aactaccagt catcttgcta atccaaactt tacaagtaat 720
tccatggaaa taaagtcagc acaggaaaat ccattcttat tcagcacaat taatcaaaaca 780
gttgaagaat taaacacaaa taaagaatct gttattgcca tttttgtacc tgcagaaaat 840
tctaaacctc cagttaattc ttttatctc gcacaaaaag aaaccacgga aatggaagac 900
acagacattg aagactcctt gtataaggat gtagactatg ggacagaagt ttacaaatc 960
gaacattctt actgcagaca agatataaat aaggaacatc tttggcagaa agtctctaag 1020
ctacattcaa agataactct tctagagtta aaagagcaac aaactctagg tagattgaag 1080
tctttggaag ctcttataag gcagctaaaag cagggaaaact ggctatctga agaaaacgtc 1140
aagattatag aaaaccattt tacaacatat gaagtcacta tgatatag 1188

```

&lt;210&gt; 165

&lt;211&gt; 669

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 165

```

atggtgaaat gctgctccgc cattggatgt gcttctcgct gcttgccaaa ttcgaagtta 60
aaaggactga catttcacgt attccccaca gatgaaaaca tcaaaaggaa atgggtatta 120
gcaatgaaaa gacttgatgt gaatgcagcc ggcatttggg agcctaaaaa aggagatgtg 180
ttgtgttcga ggcactttaa gaagacagat tttgacagaa gtgctccaaa tattaactg 240

```



```

aaacctggag tcataccttc tatctttgat tctccatata acctacaggg gaaaagagaa 300
aaacttcatt gtagaaaaaa cttcaccctc aaaaccgttc cagccactaa ctacaatcac 360
catcttggtg gtgcttcctc atgtattgaa gaattccaat cccagttcat ttttgaacat 420
agctacagtg taatggacag tccaaagaaa cttaagcata aattagatca tgtgatcggc 480
gagctagagg atacaaagga aagtctacgg aatgttttag accgagaaaa acgttttcag 540
aaatcattga ggaagacaat cagggaatta aaggatgaat gtctgatcag ccaagaaaca 600
gcaaatagac tggacacttt ctgttgggac tgttgtcagg agagcataga acaggactat 660
atttcatga                                     669

```

&lt;210&gt; 166

&lt;211&gt; 930

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 166

```

atgccgcgtc actgctccgc cgccggctgc tgcacacggg acacgcgcga gacgcgcaac 60
cgccgcatct ccttccacag acttcccaag aaggacaacc cgaggcgagg cttgtggctg 120
gccaaactgcc agcggtgga cccagcggc cagggcctgt gggacccggc atccgagtac 180
atctacttct gctccaaaca ctttgaggag gactgctttg agctggtggg aatcagtggg 240
tatcacaggc taaaggaggg ggcagtcctc accatatttg agtctttctc caagtgtgcg 300
cggacaacca agaccaaaag acacagttac ccacctggcc cccctgaagt cagccggctc 360
agacgatgca ggaagcgctg ctccgagggc cgagggccca caactccatt ttctccacct 420
ccacctgctg atgtcacctg ctttcctgtg gaagaggcct cagcacctgc cactttgccg 480
gcctcccag ctgggaggct ggagcctggc cttagcagcc ctttttcaga cctactgggc 540
cccttgggtg cccaggcaga tgaagcaggc tgcagcggcc agccttcacc agagcggcag 600
ccctcccctc tcgaaccacg gccagtctcc cctcagcgt atatgctgcg cctgccccca 660
cccgcgggag cctacatcca gaatgaacac agctaccagg tgggcagcgc cttactctgg 720
aagcggcgag ccgaggcagc ccttgatgcc cttgacaagg cccagcgcca gctgcaggcc 780
tgcaagcggc gggagcagcg gctgcggttg agactgacca agctgcagca ggagcgggca 840
cgggagaagc gggcacaggc agatgcccg cagactctga aggagcatgt gcaggacttt 900
gccatgcagc tgagcagcag catggcctga                                     930

```

&lt;210&gt; 167

&lt;211&gt; 825

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 167

```

atgcccagtg actgcagggc gccgaactgc tccaacactg cgggcccgcct ggggtgcagac 60
aaccgccctg tgagcttcta caagttccca ctgaaggatg gtccccggct gcaggcctgg 120
ctgcagcaca tgggtgtgga gcaactgggtg cccagctgcc accagcactt gtgcagcgag 180
cacttcacac cctcctgctt ccagtggcgc tggggtgtgc gctacctgcg gcctgatgca 240
gtgccctcca tcttctcccg gggaccacct gccaaagatc agcggaggac ccgaagcacc 300
cagaagccag tctgcgcgcc gcctccccta cagaagaata caccctgcc ccagagccct 360
gccatcccag tctctggccc agtgcgccta gtggtgctgg gccccacatc ggggagcccc 420
aagactgtgg ccaccatgct cctgaccccc ctggccccctg cgccaactcc tgagcgggtca 480
caacctgaag tccctgcccc acaggcccag accgggctgg gccagtgct gggagcactg 540
caacgccggg tgcggaggct gcaacgggtg caggagcggc accaggcgca gctgcaggcc 600
ctggaacggc tggcacagca gctacacggg gagagcctgc tggcacgggc acgcccgggt 660
ctgcagcgcc tgacaacagc ccagaccctt ggacctgagg aatcccaaac cttcaccatc 720
atctgtggag ggcctgacat agccatggtc cttgccagg accctgcacc tgccacagtg 780
gatgccaaagc cggagctcct ggacactcgg atccccagtg cataa                                     825

```

&lt;210&gt; 168

&lt;211&gt; 3171

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 168

```

atgacccgaa gttgctccgc agtgggctgc agcaccctg acaccgtgct cagccgggag 60
cgcggcctct ccttccacca atttccaact gataccatac agcgctcaaa atggatcagg 120

```

```

gctgttaatc gtgtggaccc cagaagcaaa aagatttggg ttccaggacc aggtgctata 180
ctgtgttcca aacatttttc agaaagtgc tttgagtcac atggcataag aagaaagctg 240
aaaaaaggag ctgtgccttc tgtttctcta tacaagattc ctcaagggtg acatcttaaa 300
ggtaaagcaa gacaaaaaat cctaaaacaa cctcttccag acaattctca agaagttgct 360
actgaggacc ataactatag tttaaagaca cctttgacga taggtgcaga gaaactggct 420
gaggtgcaac aaatgtttaca agtggtccaa aaaagactta tctccgtaaa gaactacagg 480
atgatcaaga agagaaaggg ttacgatta attgatgcac ttgtagaaga gaaactactt 540
tctgaagaaa cagagtgtct gctacgagct caattttcag attttaagtg ggagttatat 600
aattggagag aaacagatga gtactccgca gaaatgaaac aatttgcatg tacactctac 660
ttgtgcagta gcaaagtcta tgattatgta agaaagattc ttaagctgcc tcattcttcc 720
atcctcagaa cgtgggtatc caaatgccaa cccagtcag gtttcaacag caacattttt 780
tcttttcttc aacgaagagt agagaatgga gatcagctct atcaatactg ttcattgtta 840
ataaaaagta tacctctcaa gcaacagctt cagtgaggatc cttagcagta cagtttgacg 900
gggtttatgg acttttgtct tggaaaactt gatgctgatg aaacgccact tgcttcagaa 960
actgttttgt taatggcagt gggattttt ggccattgga gaacacctct tgggtatttt 1020
tttgtaaaca gagcatctgg atatttgacg gctcagctgc ttcgtctgac tattggtaaa 1080
ctgagtgcac taggaatcac agttctggct gttacatctg atgccacagc acatagtgtt 1140
cagatggcaa aagcattggg gatacatatt gatggagacg acatgaaatg tacatttcag 1200
catccttcat cttctagtca acagattgca tacttctttg actcttgcca cttgctaaga 1260
ttaataagaa atgcatttca gaattttcaa agcattcagt ttattaatgg tatagcacat 1320
tggcagcacc tcgtggagtt agtagcactg gaggaacagg aattatcaaa tatggaaaaga 1380
ataccaagta cacttgcaaa tttgaaaaat catgtactga aagtgaatag tgccacccaa 1440
ctcttttagt agagtgtagc cagtgcatta gaattttgt tatccttaga cctgccacct 1500
tttcaaaact gtattgttac catccatttt ttacgtttaa ttaacaatct gtttgacatc 1560
tttaatatga ggaactgtta tggaaagggg cttaaagggc ctctgttgcc tgaaacttac 1620
agtaaaataa accacgtgtt aattgaagcc aagactattt ttgttacatt atctgacact 1680
agcaataatc aaataattaa aggtaagcaa aaactaggat tctgtggatt tttgctcaat 1740
gctgagagct taaaatggct ctaccaaact tatgttttcc caaaggctcat gccttttct 1800
tatcttctga cttacaaact cagtcattat catctggaat tatttctaaa gatgcttagg 1860
caggtattag taacaagttc tagccctacc tgcattggcat tccagaaagc ttactataat 1920
ttggagacca gatacaaat tcaagatgaa gtttttctaa gcaaagtaag catctttgac 1980
atttcaattg ctcgaaggaa agacttgccg ctttgacag ttcaacgtca gtatgggtgc 2040
agcgttacaa agactgtctt tcacgaagag ggtatttgtc aagactggct tcattgttca 2100
ctaagtgagg cattactaga cctgtcagat cataggcgaa atctcatctg ttatgctgg 2160
tatgttgcaa acaagttatc agctcttcta actgttgagg actgcatcac tgcactgtat 2220
gcatcgatc tcaaagcctc taaaattggg tcaactattt ttgttaaaaa gaagaatgg 2280
ttgcattttc cttcagaaag tctgtgtcgg gtcataaata tttgtgagcg agttgtaaga 2340
acctattcaa gaatggcaat ttttgaacta gtttctaaac aaagggaatt gtatcttcaa 2400
cagaaaatat tatgtgagct ttctgggcat attgatcttt ttgtagatgt gaataagcat 2460
ctctttgatg gagaagtgtg tgccatcaat cactttgtca agttgctaaa ggatataata 2520
atctgtttct taaatatcag agctaaaaat gttgcacaga atcctttaaa acatcattca 2580
gagagaactg atatgaaaac tttatcaagg aaacactggc cacctgtaca ggattataaa 2640
tgttcaagtt ttgctaatac cagtagtaaa ttcaggcatt tgctaagtaa cgatggatat 2700
ccattcaaat gagagaccta aaatatatta acattttaat taagaatact tgatcaacat 2760
tttttgaagt tcaatttacc atattttata aattgcgcac tctgcacagt ggacaagttt 2820
gcaattctga cttattaaaa tttcaaattc tgcatatcac aaaatctcct tatacttttg 2880
gtatggcttg cagcatttat gagttttcca aaatatagaa agcagtaggt cagtaggagc 2940
aaactagcca acaggtactg tctttgaatt tactactgta agactaagca gtgttactgg 3000
acacagtttt aacttgttca atctgcttca aaaacaagaa aaacaacaac tatgagttat 3060
caaaatatgg actccattta tgactagact acatttctga aagatctttg gtttacgatt 3120
cttaagaata ttgacaatac ctataaaact ttgaagataa cttttactta a 3171

```

&lt;210&gt; 169

&lt;211&gt; 774

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 169

```

atgccggccc gttgtgtggc cgcccactgc ggcaacacca ccaagtctgg gaagtcgctg 60
ttccgctttc ccaaggaccg ggccgtgcgg ctgctctggg accgcttcgt gcggggttgc 120
cgcgccgact ggtacggagg caatgaccgc tcggctcatc gctctgacca ctttgcccca 180

```

```

gacctgtttg acgtctcttc ggttatccag aagaacctgc gcttctccca gcgcctgagg 240
ctgggtggcag gcgcctgccc caccctgcac cgggtgcccg ccccggcacc taagagggga 300
gaggagggag accaagcagg ccgcctggac acgcgaggag agctccaggc agccaggcat 360
tctgaggctg ccccgaggcc agtctcctgt acacgcccc gagctgggaa gcaggctgca 420
gcttcacaga ttacgtgtga aaatgaactt gtgcaaacc aaccccatgc tgataatcca 480
tctaatactg tcacttcagt acctactcac tgtgaagaag gccagtgca taaaagtaca 540
caaatttctt tgaaaaggcc ccgtcacccg agtgtgggta ttcaagccaa agtgaaagcg 600
tttgaaaaaa gactgtgtaa tgcaactact cagacagagg aattgtggtc tagaacttcc 660
tctctctttg acatttactc cagtgtattca gaaacagata cagactggga tatcaagagt 720
gaacagagtg atttgtctta tatggctgta caggtgaaag aagaaacatg ttaa 774

```

&lt;210&gt; 170

&lt;211&gt; 945

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 170

```

atgcctggct ttacgtgctg cgtgccaggc tgctacaaca actcgcaccg ggacaaggcg 60
ctgcacttct acacgtttcc aaaggacgct gagttgcccg gcctctggct caagaacgtg 120
tcgcgtgccg gcgtcagtggt gtgcttctcc acctccagc ccaccacagg ccaccgtctc 180
tgacgcgttc acttccaggg cggccgcaag acctacacgg tacgcgtccc caccatcttc 240
ccgctgcgcg gcgtcaatga gcgcaaagta gcgcgcagac ccgctggggc cgcggccgcc 300
cgccgcaggc agcagcagca acagcagcag cagcagcaac agcagcaaca gcagcagcag 360
cagcaacagc agcagcagca gcagcagcag cagcagtcct caccctctgc ctccactgcc 420
cagactgccc agctgcagcc gaacctggta tctgcttccg cggccgtgct tctcaccctt 480
caggccactg tagacagcag tcaggtcccg ggatccgtac agccggcgcc catcactccc 540
actggagaag acgtgaagcc catcgatctc acagtgaag tggagtttgc agccgcagag 600
ggcgcagccc ctgcggccgc cgcgtcggag ttacaggctg ctaccgcagg gctggaggct 660
gccgagtgcc ctatgggccc ccagttgggt gtggtagggg aagagggtct ccctgatact 720
ggctccgacc attcgtactc cttgtcgtca ggcaccacgg aggaggagct cctgcgcaag 780
ctgaatgagc agcgggacat cctggctctg atggaagtga agatgaaaga gatgaaaggc 840
agcattcgcc acctgcgtct cactgaggcc aagctgcgcg aagaactgcg tgagaaggat 900
cggctgcttg ccatggctgt catccgcaag aagcacgga tgtga 945

```

&lt;210&gt; 171

&lt;211&gt; 2286

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 171

```

atgccgaact tctgcgtgc ccccaactgc acgcggaaga gcacgcagtc cgacttggcc 60
ttcttcaggt tcccgcgga cctgccaga tgccagaagt ggggtggagaa ctgtaggaga 120
gcagacttag aagataaaac acctgatcag cttaaataaac attatcgatt atgtgccaaa 180
cattttgaga cctctatgat ctgtagaact agtccttata ggacagttct tcgagataat 240
gcaataccaa caatatttga tcttaccagt catttgaaca acccacatag tagacacaga 300
aaacgaataa aagaactgag tgaagatgaa atcaggacac tgaaacagaa aaaaattgat 360
gaaacttctg agcaggaaca aaaacataaa gaaaccaaca atagcaatgc tcagaacccc 420
agcgaagaag aggggtgaagg gcaagatgag gacattttac ctctaaccct tgaagagaag 480
gaaaacaaag aatacctaaa atctctattt gaaatcttga ttctgatggg aaagcaaaac 540
atacctctgg atggacatga ggctgatgaa atcccagaag gtctctttac tccagataac 600
tttcaggcac tgctggagtg tcggataaat tctggtgaag aggttctgag aaagcggttt 660
gagacaacag cagttaacac gttgttttgt tcaaaaacac agcagaggca gatgctagag 720
atctgtgaga gctgtattcg agaagaaact ctcagggaag tgagagactc acacttcttt 780
tccattatca ctgacgatgt agtggacata gcaggggaag agcacctacc tgtgttgggtg 840
aggtttgttg atgaatctca taacctaaaga gaggaattta taggcttccg gccttatgaa 900
gccgatgcag aaattttggc tgtgaaattt cacactatga taactgagaa gtggggatta 960
aatatggagt attgtcgtgg ccaggcttac attgtctcta gtggattttt ttccaaaatg 1020
aaagtgtgtg cttctagact tttagagaaa tatccccaag ctatctacac actctgctct 1080
tctgtgcct taaatatgtg gttggcaaaa tcagtacctg ttatgggagt atctgttgca 1140
ttaggaacaa ttgaggaagt ttgttctttt ttccatcgat caccacaact gcttttagaa 1200
cttgacaacg taatttctgt tctttttcag aacagtaaaag aaaggggtaa agaactgaag 1260

```

```

gaaatctgcc attctcagtg gacaggcagg catgatgctt ttgaaatddd agtggaaactc 1320
ctgcaagcac ttgttttatg tttagatggg ataaatagtg acacaaatat tagatggaat 1380
aactatatag ctggccgagc atttgtactc tgcagtgcag tgtcagattt tgatttcatt 1440
gttactattg ttgttcttaa aaatgtccta tcttttacia gagcctttgg gaaaaacctc 1500
caggggcaaa cctctgatgt cttctttgcg gccggtagct tgactgcagt actgcattca 1560
ctcaacgaag tgatggaaaa tattgaagtt tatcatgaat tttggtttga ggaagccaca 1620
aatttgga ccaaacttga tattcaaatt aaactccctg ggaaattccg cagagctcac 1680
cagggtaact tggaatctca gctaacctct gagagttact ataaagaaac cctaagtgtc 1740
ccaacagtg agcacattat tcaggaactt aaagatatat tctcagaaca gcacctcaa 1800
gctcttaaat gcttatctct ggtacctca gtcattggac aactcaaatt caatacgtcg 1860
gaggaacacc atgctgacat gtatagaagt gacttaccac atcctgacac gctgtcagct 1920
gagcttcatt gttggagaat caaatggaaa cacaggggga aagatataga gcttccgtcc 1980
accatctatg aagccctcca cctgcctgac atcaagtttt ttcctaattg gtatgcattg 2040
ctgaagggtc tgtgtattct tcctgtgatg aaggttgaga atgagcggta tgaaaatgga 2100
cgaaagcgtc ttaaagcata tttgaggaac actttgacag accaaagggtc aagtaacttg 2160
gctttgctta acataaattt tgatataaaa cagcactgg atttaattgg ggacacatat 2220
attaaactct atacaagtaa gtcagagctt cctacagata attccgaaac tgtggaaaat 2280
acctaa
2286

```

&lt;210&gt; 172

&lt;211&gt; 633

&lt;212&gt; DNA

&lt;213&gt; Mus musculus

&lt;400&gt; 172

```

atggtgcagt cctgctccgc ctacggctgc aagaaccgct acgacaagga caagcccgtc 60
tccttcacaca agtttccctct tactcgcccc agcctttgta agcagtggga ggcagctgtt 120
aaaaggaaaa acttcaagcc caccaagtac agcagcatct gctcggagca cttcaccctg 180
gactgcttta agagggagtg caacaacaag ctactgaagg agaacgctgt gccacacata 240
tttctctata tcgagccaca tgagaagaag gaagacctgg aatcccaaga acagctcccc 300
tctccttcac cccccgcttc ccagggtgat gctgctattg ggctgctaatt gccccctctg 360
cagacccttg ataacctgtc ggttttctgt gaccacaatt acactgtgga ggatacgtg 420
caccagagga agaggatcct gcagctggag cagcaggtgg agaaactcag gaagaagctc 480
aagacggccc agcagcgggtg ccggcggcag gagaggcagc tcgagaagct caaggaagtc 540
gtccactttc agagagagaa ggacgacgag tccgagaggg gctacgtgat cctaccaaatt 600
gactactttg aaattgttga agttccagca tga
633

```

&lt;210&gt; 173

&lt;211&gt; 654

&lt;212&gt; DNA

&lt;213&gt; Mus musculus

&lt;400&gt; 173

```

atgccgacca attgcgccgc ggccggctgt gctgctacct acaacaagca cattaacatc 60
agcttcacaca ggtttccttt ggatcctaata agaagaaaag aatgggttcg cctgggttagg 120
cgcaaaaatt ttgtgccagg aaaacacact tttctttgct caaagcactt tgaagcctcc 180
tgttttgatc taacaggaca aaccggacga cttaaaatgg atgctgttcc aaccattttt 240
gattttttgta cccatataaaa gtctctgaaa ctcaagtcaa ggaatcttct gaagacaaac 300
aacagttttc ctccaactgg accatgtaatt ttaaagctga acggcagtc gcaagtactg 360
cttgaacaca gttatgcctt taggaacctt atggaggcga aaaaaaggat aattaaacta 420
gaaaaggaaa tagcaagcct gagaaaaaaa atgaaaactt gcctgcaaag agaacgcaga 480
gcaactcgaa ggtggatcaa agccacgtgc tttgtgaaga gcttagaagc aagtaacatg 540
ctacctaagg gcatctcaga acagatttta ccaactgcct taagcaatct tctctggaa 600
gatttaaaaa gtcttgaaca agatcaacaa gataaaacag taccatttct ctaa
654

```

&lt;210&gt; 174

&lt;211&gt; 657

&lt;212&gt; DNA

&lt;213&gt; Mus musculus

&lt;400&gt; 174

```

atgccgaagt cttgcgcggc ccggcaatgc tgcaaccgct acagcagccg caggaagcag 60
ctcaccttcc accggttccc ctccagccgc ccggagctgt tgagggagtg ggtgctcaac 120
atcggccggg ctgacttcaa gcctaagcag cacacagtca tctgctcgga acacttcaga 180
cccgagtgtc tcagcgcctt tgggaaccgc aagaacctga aacacaatgc tgtgcccacg 240
gtgttcgctt ttcagaacct cacagaggctc tgccctgagg tgggggctgg tggggacagc 300
tcagggagga acatggacac cacactggaa gaacttcagc ctccaacccc ggaaggcccc 360
gtgcagcagg tcttaccaga tcgagaagca atggaggcca cggaggccgc tggcctgcct 420
gccagccctc tggggttgaa gagggccctt ccgggacagc cgtctgatca cagttatgcc 480
ctttcggact tggataccct caaaaaaaaa ctctttctca cactgaagga aaacaagagg 540
cttcggaagc ggctgaaagc ccagaggctg ctgttgcgga ggacatgtgg ccgcctgaga 600
gcctacagag agggacagcc gggacctcgg gccagacggc cggcacaggg aagctga 657

```

&lt;210&gt; 175

&lt;211&gt; 558

&lt;212&gt; DNA

&lt;213&gt; Mus musculus

&lt;400&gt; 175

```

atactgcaag catttggaag cctaaaaaaaa ggagatgtgc tgtgttcaag acacttcaag 60
aagacagact ttgacagaag cactctaaac actaagctga aggcaggagc catcccttct 120
atctttgaat gtccatatca cttacaggag aaaagagaaa aacttctactg tagaaaaaac 180
ttccttctca aaacccttcc catcacccac catggccgcc agcttgttgg tgcctcctgc 240
attgaagaat tcgaacccca gttcattttt gaacatagct acagtgttat ggacagccca 300
aagaagctta agcataagct agaccgtgtg atcatcgagc tggagaatac caaggaaaagc 360
ctacggaatg ttttagcccg agaaaaacac tttcaaaagt cactgaggaa gacaatcatg 420
gaactaaagg atgaaagtct gatcagccag gaaacagcca atagtctggg tgctttctgt 480
tgggagtgtc atcatgaaag cacagcagga ggctgtagtt gtgaagtcac ttcttatatg 540
cttcatctgc agttgaca 558

```

&lt;210&gt; 176

&lt;211&gt; 1719

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 176

```

ctttccgcgc ggcggaagag cgcgcgccag cttcggcaca cttgggagcc ggatcccagc 60
cctacgcctc gtcccctaca agctcctcca agccccgccg gctgctgtgg gagcggcggc 120
cgctctctcc tggaggtcgt ctctggcat cctcggggcc gcaggaagga agaggaggca 180
gcggccggag ccctgggtgg cggcctgagg tgagagccc accggccctt ttgggaatat 240
ggcgaccggt ggctaccgga ccagcagcgg cctcggcgcc agcaccacag acttcttggg 300
ggagtggaag gcgaaacgcg agaagatgcg cgccaagcag aacccccccg gcccgggccc 360
cccgggaggg ggcagcaagc acgcccgtgg gaagcccccc gcgggggctc tgggcacccc 420
ggcgccgcgc gctgccaacg agtcaacaa caacctccc ggcgccgcgc cggccgcacc 480
tgccgtcccc ggtcccgggg gcgtgaactg cgcggtcggc tccgccatgc tgacgcgggc 540
gcccccgccc gcgcccgcgc ggcggtcgga ggacgagccc ccagccgcct ctgcctcggc 600
tgacccgccc cccagcgtg acgaggagga gccggacggc gtcccagaga agggcaagag 660
ctcgggcccc agtgccagga aaggcaaggg gcagatcgag aagaggaaagc tgcgggagaa 720
gcggcgctcc accggcggtg tcaacatccc tgccgcagag tgcttagatg agtacgaaga 780
tgatgaagca ggcgcaaaag agcggaaaac agaagatgca attacacaac agaacactat 840
tcagaatgaa gctgtaaact tactagatcc aggcagttcc tatctgctac aggagccacc 900
tagaacagtt tcaggcagat ataaaagcac aaccagtgtc tctgaagaag atgtctcaag 960
tagatattct cgaacagata gaagtgggtt ccctagatat aacagggatg caaatgtttc 1020
aggtactctg gtttcaagta gcacactgga aaagaaaatt gaagatcttg aaaaggaagt 1080
agtaacagaa agacaagaaa acctaagact tgtgagactg atgcaagata aagaggaaat 1140
gatttgaaaa ctcaaagaag aaattgattt attaaataga gacctagatg acatagaaga 1200
tgaaaatgaa cagctaaagc aggaaaataa aactcttttg aaagtgtgg gtcagctgac 1260
caggtagagg attcaagact caatgtggaa aaaatatttt aaactactga ttgaatgtta 1320
atggtcaatg ctagcacaat attcctatgc tgcaatacat taaaataact aagcaagtat 1380
atattttct agcaaacaga tgtttgtttt caaaatactt ctttttcatt attggtttta 1440
aaaaagcatt atccttttat ctcacaaata agtaatatct ttcagttatt aaatgataga 1500
taatgccttt ttggttttgt gtggtattca actaatacat ggtttaaagt cacagccgtt 1560

```

```

tgaatatatt ttatcttgggt agtacatddd ctcccttagg aatatacata gtctttgttt 1620
acatgagttc caatactttt gggatgttac cctcacatgt ccctatactg atgtgtgcca 1680
ccttttatgt gttgatgact cactcataag gttttggtc 1719

```

<210> 177  
 <211> 878  
 <212> DNA  
 <213> Homo sapiens

```

<400> 177
atcccagccc acgcacagac ccccaacttg cagctgcccc cctcacccctc agctctggcc 60
tcttactcac cctctaccac agacatggct cagtcactgg ctctgagcct ccttatectg 120
gttctggcct ttggcatccc caggacccaa ggcagtgatg gaggggctca ggactgttgc 180
ctcaagtaca gccaaaggaa gattcccggc aaggttgctc gcagctaccg gaagcaggaa 240
ccaagcttag gctgtcccat cccagctatc ctgttcttgc cccgcaagcg ctctcaggca 300
gagctatgtg cagacccaaa ggagctctgg gtgcagcagc tgatgcagca tctggacaag 360
acaccatccc cacagaaacc agcccagggc tgcaggaagg acaggggggc ctccaagact 420
ggcaagaaag gaaagggtc caaaggctgc aagaggactg agcggtcaca gaccctaaa 480
gggccatagc ccagttagca gcctggagcc ctggagaccc caccagcctc accagcgctt 540
gaagcctgaa cccaagatgc aagaaggagg ctatgtcagc gggccctgga gcagccaccc 600
catgtgggcc ttgccacact ctttctcttg ctttaaccac cccatctgca ttcccagctc 660
tacctgcat ggctgagctg cccacagcag gccagggtcca gagagaccga ggaggagag 720
tctcccaggg agcatgagag gaggcagcag gactgtcccc ttgaaggaga atcatcagga 780
ccctggacct gatacggctc cccagtacac cccacctctt ccttgtaaata atgatttata 840
cctaactgaa taaaagctg ttctgtcttc ccacccaa 878

```

<210> 178  
 <211> 34  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Interferon gamma homology motif of THAP1

```

<400> 178
Asn Tyr Thr Val Glu Asp Thr Met His Gln Arg Lys Arg Ile His Gln
1           5           10           15
Leu Glu Gln Gln Val Glu Lys Leu Arg Lys Lys Leu Lys Thr Ala Gln
20           25           30
Gln Arg

```

<210> 179  
 <211> 20  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Nuclear localization sequence of THAP1

```

<400> 179
Arg Lys Arg Ile His Gln Leu Glu Gln Gln Val Glu Lys Leu Arg Lys
1           5           10           15
Lys Leu Lys Thr
20

```

<210> 180  
 <211> 38  
 <212> PRT

<213> Artificial Sequence

<220>

<223> Consensus sequence for PAR4 binding domain of THAP

<221> UNSURE

<222> 3-16, 19, 23, 25-35

<223> Xaa = any of the twenty amino acids

<221> VARIANT

<222> 37

<223> Xaa = Arg or Lys

<400> 180

```

Leu Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 1           5           10           15
Gln Arg Xaa Arg Arg Gln Xaa Arg Xaa Xaa Xaa Xaa Xaa Xaa Xaa
          20           25           30
Xaa Xaa Xaa Gln Arg Glu
          35

```

<210> 181

<211> 50

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 181

gaattcggcc attatggcct gcaggatccg gccgcctcgg cccaggatcc 50

<210> 182

<211> 111

<212> PRT

<213> Homo sapiens

<220>

<400> 182

```

Ser Asp Gly Gly Ala Gln Asp Cys Cys Leu Lys Tyr Ser Gln Arg Lys
 1           5           10           15
Ile Pro Ala Lys Val Val Arg Ser Tyr Arg Lys Gln Glu Pro Ser Leu
          20           25           30
Gly Cys Ser Ile Pro Ala Ile Leu Phe Leu Pro Arg Lys Arg Ser Gln
          35           40           45
Ala Glu Leu Cys Ala Asp Pro Lys Glu Leu Trp Val Gln Gln Leu Met
          50           55           60
Gln His Leu Asp Lys Thr Pro Ser Pro Gln Lys Pro Ala Gln Gly Cys
          65           70           75           80
Arg Lys Asp Arg Gly Ala Ser Lys Thr Gly Lys Lys Gly Lys Gly Ser
          85           90           95
Lys Gly Cys Lys Arg Thr Glu Arg Ser Gln Thr Pro Lys Gly Pro
          100          105          110

```

<210> 183

<211> 37

<212> DNA

<213> Artificial Sequence

<220>  
<223> Primer

<400> 183  
gcgggatccg tagtgatgga ggggctcagg actgttg 37

<210> 184  
<211> 35  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 184  
gcgggatccc tatggccctt taggggtctg tgacc 35

<210> 185  
<211> 33  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 185  
ccgaattcag gatggtgcag tctgtctcg cct 33

<210> 186  
<211> 39  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 186  
cgcggtacct gctggtactt caactatttc aaagtagtc 39

<210> 187  
<211> 33  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 187  
ccgaattcag gatggtgcag tctgtctcg cct 33

<210> 188  
<211> 39  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 188  
cgcggtacct gctggtactt caactatttc aaagtagtc 39



<210> 189  
<211> 33  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 189  
gcggaattca tggcgaccgg tggctaccgg acc

33

<210> 190  
<211> 35  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 190  
gcgggatccc tctacctggt cagctgaccc acaac

35

<210> 191  
<211> 33  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 191  
ccgaattcag gatggtgcag tcttgcctcg cct

33

<210> 192  
<211> 39  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 192  
cgcgatcct gctggtactt caactatttc aaagtagtc

39

<210> 193  
<211> 46  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 193  
cgcgattcgc ccatcatggg gttccctaga tataacaggg atgcaa

46

<210> 194  
<211> 37  
<212> DNA  
<213> Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 194

gccggatccg ggttcctag atataacagg gatgcaa

37

&lt;210&gt; 195

&lt;211&gt; 37

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 195

gcgctctaga gccatcatgg aggagcagaa gctgatc

37

&lt;210&gt; 196

&lt;211&gt; 37

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 196

cttgcggccg cctctacctg gtcagctgac ccacaac

37

&lt;210&gt; 197

&lt;211&gt; 37

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 197

gcggaattca aagaagatct tctggagcca caggaac

37

&lt;210&gt; 198

&lt;211&gt; 39

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 198

cgcggatcct gctgtactt caactatttc aaagtagtc

39

&lt;210&gt; 199

&lt;211&gt; 35

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 199

gcggaattca tgccgcctct tcagaccctt gttaa

35

<210> 200  
<211> 36  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 200  
gcggaattca tgcaccagcg gaaaaggatt catcag 36

<210> 201  
<211> 33  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 201  
ccgaattcag gatggtgcag tctgtctcgc cct 33

<210> 202  
<211> 39  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 202  
gcgggatccc ttgtcatgtg gctcagtaca aagaaatat 39

<210> 203  
<211> 34  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 203  
cgggatcctg tgcggtcttg agcttctttc tgag 34

<210> 204  
<211> 36  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 204  
gcgggatccg tcgtctttct ctttctggaa gtgaac 36

<210> 205  
<211> 36  
<212> PRT  
<213> Artificial Sequence

<220>

<223> Consensus sequence for PAR4 binding domain of THAP

<221> UNSURE

<222> 3-14, 17, 21, 23-33, 35

<223> Xaa = any of the twenty amino acids

<400> 205

Leu	Glu	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Gln	Arg
1			5				10						15		
Xaa	Arg	Arg	Gln	Xaa	Arg	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa
			20				25						30		
Xaa	Gln	Xaa	Glu												
			35												

<210> 206

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 206

ccgcacagca gcgatgcgct gctcaagaac ggcagcttg 39

<210> 207

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 207

caagctgccg ttcttgagca gcgcacgcgct gctgtgcgg 39

<210> 208

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 208

gctcaagacc gcacagcaag aacggcagct tg 32

<210> 209

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 209

caagctgccg ttcttgctgt gcggtcttga gc 32

<210> 210

<211> 36

<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 210  
gcgggatccc taaattagaa aggggtgggg gtagcc 36

<210> 211  
<211> 32  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 211  
gcggaattca tggagcctgc acccgcccga tc 32

<210> 212  
<211> 37  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 212  
gcggaattca aagaagatct tctggagcca caggaac 37

<210> 213  
<211> 39  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 213  
cgcggtacct gctgtactt caactatttc aaagtagtc 39

<210> 214  
<211> 33  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 214  
cgcggtaccg tgcagtctg ctccgcctac ggc 33

<210> 215  
<211> 39  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 215  
ccgaattctt atgctggtac ttcaactatt tcaaagtag  
39

<210> 216  
<211> 33  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 216  
gccgaattca tgccgaactt ctgctgctgcc ccc  
33

<210> 217  
<211> 40  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 217  
cgcggtatcct taggttattt tccacagttt cggaattatc  
40

<210> 218  
<211> 39  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 218  
gcgctgcagc aagctaaatt taaatgaagg tactcttgg  
39

<210> 219  
<211> 35  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 219  
gcgagatctg ggaaatgccg accaattgctg ctgctg  
35

<210> 220  
<211> 35  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 220  
agaggatcct tagctctgct gctctggccc aagtc  
35

<210> 221  
<211> 32  
<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 221

agagaattca tgccgaagtc gtgcgcggcc cg

32

<210> 222

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 222

gcggaattca tgccgcgtca ctgctccgcc gc

32

<210> 223

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 223

gcgggatcct caggccatgc tgctgctcag ctgc

34

<210> 224

<211> 38

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 224

gcgagatctc gatggtgaaa tgctgctccg ccattgga

38

<210> 225

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 225

gcgggatcct catgaaatat agtctgttc tatgctctc

39

<210> 226

<211> 35

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 226

gcgagatctc gatgcccaag tactgcaggg cgccg 35

<210> 227  
<211> 37  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 227  
gcggaattct tatgcactgg ggatccgagt gtccagg 37

<210> 228  
<211> 32  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 228  
gcggaattca tgccggcccg ttgtgtggc gc 32

<210> 229  
<211> 39  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 229  
gcgggatacct taacatgttt cttctttcac ctgtacagc 39

<210> 230  
<211> 36  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 230  
gcgagatctc gatgcctggc ttacgtgct gcgtgc 36

<210> 231  
<211> 36  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 231  
gcggaattct cacattccgt gttcttgcg gatgac 36

<210> 232  
<211> 33  
<212> DNA  
<213> Artificial Sequence



&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 232

ccgaattcag gatggtgcag tcttgctccg cct

33

&lt;210&gt; 233

&lt;211&gt; 39

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 233

cgcggtatcct gctggtactt caactatttc aaagtagtc

39

&lt;210&gt; 234

&lt;211&gt; 37

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 234

gcgctctaga gccatcatgg aggagcagaa gctgac

37

&lt;210&gt; 235

&lt;211&gt; 41

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 235

gcgctctaga ttatgctggt acttcaacta tttcaaagta g

41

&lt;210&gt; 236

&lt;211&gt; 33

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 236

cgcggtatccg tgcagtcctg ctccgcctac ggc

33

&lt;210&gt; 237

&lt;211&gt; 39

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 237

cgcggtatcct gctggtactt caactatttc aaagtagtc

39

<210> 238  
<211> 37  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Primer  
  
<400> 238  
gccggatccg ggttccttag atataacagg gatgcaa 37  
  
<210> 239  
<211> 35  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Primer  
  
<400> 239  
gcgggatccc tctacctggt cagctgaccc acaac 35  
  
<210> 240  
<211> 35  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Primer  
  
<400> 240  
gcgggatcca gtgatggagg ggctcaggac tgttg 35  
  
<210> 241  
<211> 35  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Primer  
  
<400> 241  
gcgggatccc tatggccctt taggggtctg tgacc 35  
  
<210> 242  
<211> 33  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Primer  
  
<400> 242  
gcgcataatgg tgcagtcctg ctccgcctac ggc 33  
  
<210> 243  
<211> 36  
<212> DNA  
<213> Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 243

gcgctcgagt ttcttgatcat gtggctcagt acaaag

36

&lt;210&gt; 244

&lt;211&gt; 62

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Oligonucleotide

&lt;221&gt; unsure

&lt;222&gt; 21-45

&lt;223&gt; n = any of the four nucleotides

&lt;400&gt; 244

tgggcactat ttatatcaac nnnnnnnnnn nnnnnnnnnn nnnnnaatgt cggttggtggc 60  
cc 62

&lt;210&gt; 245

&lt;211&gt; 30

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 245

accgcaagct tgggcactat ttatatcaac

30

&lt;210&gt; 246

&lt;211&gt; 24

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 246

ggtctagagg gccaccaacg catt

24

&lt;210&gt; 247

&lt;211&gt; 2173

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 247

gacgggcat ggctgtggc cttctgctaa tgcaaacaac aaaacgggca cactagtcac 60  
ccccgagga gccaccatc actgtaactg ttggccaaag ctacaaaaga agcgagggaa 120  
tccaaccgag cgcagcgaca ctgagaacag cttcccctgc cttctgcggc ggcagaagtg 180  
aagtgcctga ggaccggaag gatggtgcag tcctgctccg cctacggctg caagaaccgc 240  
tacgacaagg acaagcccgt ttctttccac aagtttctc ttactcgacc cagtctttgt 300  
aaagaatggg aggcagctgt cagaagaaaa aactttaaac ccaccaagta tagcagtatt 360  
tggtcagagc actttactcc agactgcttt aagagagagt gcaacaacaa gttactgaaa 420  
gagaatgctg tgcccacaat atttctttgt actgagccac atgacaagaa agaagatctt 480  
ctggagccac aggaacagct tccccacct cttttaccgc ctctgtttc ccaggttgat 540  
gctgctattg gattactaat gccgcctctt cagaccctg ttaatctctc agttttctgt 600  
gaccacaact atactgtgga ggatacaatg caccagcgga aaaggattca tcagctagaa 660

```

cagcaagttg aaaaactcag aaagaagctc aagaccgcac agcagcgatg cagaaggcaa 720
gaacggcagc ttgaaaaatt aaaggaggtt gttcacttcc agaaagagaa agacgacgta 780
tcagaaaagag gttatgtgat tctaccaaat gactactttg aaatagttga agtaccagca 840
taaaaaaatg aaatgtgtat tgattttctaa tggggcaata ccacatatcc tcctctagcc 900
tgtaaaggag tttcatttaa aaaaataaca tttgattact tatataaaaa cagttcagaa 960
tattttttta aaaaaaattc tatatatact gtaaaattat aaattttttt gtttgtaatt 1020
tcagggtttt tacattttta caaaatattt taaaagttat aaactaacct cagacctcta 1080
atgtaagttg gtttcaagat tggggatttt ggggtttttt ttttagtatt tatagaaata 1140
atgtaaaaaa aaaaagtaaa gagaatgaga acagtgtggt aaaagggtga tttcagttta 1200
aaacttaaaa ttagtactgt tttattgaga gaatttagtt atatttttaa tcagaagtat 1260
gggtcagatc atgggacata acttcttaga atatataat acatatgtac atatttctcat 1320
atgtaaaagtc acaagggttca tttatctttc tgaatcagtt atcaaagata aattggcaag 1380
tcagtactta agaaaaaaga tttgattatc atcacagcag aaaaaagtca ttgcataatc 1440
gatcaataac ttcagattct aagagtggat tttttttttt tacatgggct cctatttttt 1500
cccctactgt cttgcattat aaaattagaa gtgtattttc agtggaaagaa acattttttca 1560
ataaataaag taaggcattg tcatcaatga agtaattaaa actgggacct gatctatgat 1620
acgctttttt ctttcattac accctagctg aaggacatcc cagttcccca gctgtagtta 1680
tgtatctgcc ttcaagttc tgcacaaatgt gctgtgttag tagagttaga tttgtatcat 1740
atgataatct tgcacttgac tgagttggga caaggcttca cataaaaaat tatttcttca 1800
cttttaacac aagttagaaa ttatatccca tttagttaaa tgcgtgattt atattcagaa 1860
caacctacta tgtagcgttt attttactga atgtggagat ttaaacactg aggtttctgt 1920
tcaaactgtg agttctgttc tttgtgagaa attttacata tattggaagt gaaaatatgt 1980
tctgagtaaa caaatattgc tatgggagtt atctttttag atttagaata actgttccaa 2040
tgataattat tacttttata tttcaaagta cactaagatc gttgaagagc aatagaacct 2100
ttaagacagt attaaagggtg tgaaacaaaa aaaaaaaaaa taaaaaaaaa aaaaaaaaaa 2160
aaaaaaaaaa aaa

```

2173

&lt;210&gt; 248

&lt;211&gt; 1302

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 248

```

aattgctctg aggaccgctg ccaaagaaac gcagtagatc cgctccctct tggggggcggg 60
gagaaagaac gggttgtgtc cgccatgttg gtgaagtcaa gcgaaggcga ctagagctcc 120
aggaggcca gttctgtggg ctctagtctg ccatattaat aaagagaaag ggaaggctga 180
cgtctctcg cctccgccc caccatacaca ccccttcttc ccactccgct ctcacgacta 240
agctctcaag attaaggcac gcctgcctcg attgtccagc ctctgccaga agaaagctta 300
gcagccagcg cctcagtaga gacctaaagg cgctgaatga gtgggaaagg gaaatgccga 360
ccaattgctg tgcggcgggc tgtgccacta cctacaacaa gcacattaac atcagcttcc 420
acaggtttcc tttggatcct aaaagaagaa aagaatgggt tgcctgggt aggcgcacaaa 480
attttgtgcc aggaacacac acttttcttt gttcaaagca ctttgaagcc tctgtttttg 540
acctaacagg acaaaactga cgacttaaaa tggatgctgt tccaaccatt tttgattttt 600
gtacccatat aaagtctatg aaactcaagt caaggaatct tttgaagaaa aacaacagtt 660
gttctccagc tggaccatct aatttaaaat caaacattag tagtcagcaa gtactacttg 720
aacacagcta tgccttttag aatcctatgg aggcaaaaaa gaggatcatt aaactggaaa 780
aagaaatagc aagcttaaga agaaaaatga aaacttgcct acaaaaaggaa cgcagagcaa 840
ctcgaagatg gatcaaagcc acgtgttttg taaagaattt agaagcaaat agtgtattac 900
ctaaaggtag atcagaacac atgttaccaa ctgccttaag cagtcttctt ttggaagatt 960
ttaagatcct tgaacaagat caacaagata aaacactgct aagtctaaat ctaaaacaga 1020
ccaagagtac cttcatttaa atttagcttg cacagagctt gatgcctatc cttcattctt 1080
ttcagaagta aagataatta tggcacttat gccaaaattc attatttaat aaagttttac 1140
ttgaagtaac attactgaat ttgtgaagac ttgattacaa aagaataaaa aacttcatat 1200
ggaaatttta tttgaaaatg agtggaagtg ccttacatta gaattacgga cttaaaaatt 1260
ttgctaataa atttgtgtatt tgaaaaaaaa aaaaaaaaaa aa

```

1302

&lt;210&gt; 249

&lt;211&gt; 1995

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 249

```

ccagtgcagt cagaggagtc cagacctatt cacaattcaa agccctaaaa acactgaggg 60
gttggccgtt ggtttccagt tgtccaagcc tgtgagtgcc tatgcgtcct ggttgggtgc 120
tcaaagcaag gaggtgaaag gcgaccagca ttggcgaatg gggtaagact tgcacaggcc 180
caaggctagg agttgggggt tcgggcctga attggggccc ggagcaccctc tttacgtggc 240
gccccgggtc ccgtccgacc ctggggagac gcgggtggct gggatggcag gatgagcgcg 300
ccctggaggg gagccaggcc cgtcaccacc tcccagcggc cccgcccctc cccgcaggtc 360
cctcccctct ccgcaggccc cgccgcgcgc gccatctttg ttggggggcag ccaggccttg 420
ctcgagatgc cgaagtcgtg cgcggcccg cagtgtctga accgctacag cagccgcagg 480
aagcagctca ccttccaccg gtttcggtt agccgcccgg agctgctgaa ggaatgggtg 540
ctgaacatcg gccggggcaa cttcaagccc aagcagcaca cggtcactct ctccgagcac 600
ttccggccag agtgcttcag cgcttttggg aaccgcaaga acctaaagca caatgccgtg 660
cccacggtgt tcgcttttca ggaccccaca cagcaggtga gggagaacac agaccctgcc 720
agtgaagag gaaatgccag ctcttctcag aaagaaaagg tcctccctga ggcggggggc 780
ggagaggaca gtcctgggag aaacatggac actgcacttg aagagcttca gttgccccca 840
aatgccgaag gccacgtaaa acaggtctcg ccacggaggc cgcaagcaac agaggctgtt 900
ggccggccga ctggccctgc aggcctgaga aggaccccca acaagcagcc atctgatcac 960
agctatgccc ttttggactt agattccctg aagaaaaaac tcttccctac tctgaaggaa 1020
aatgaaaagc tccggaagcg cttgcaggcc cagaggcttg tgatgcgaag gatgtccagc 1080
cgctccgtg cttgcaaagg gcaccaggga ctccaggcca gacttgggac agagcagcag 1140
agctgagccc cacaggctcc ggacgcagag gtggcagtg caccagggcc ggcagagctt 1200
tgagctctg gctgtggaca tttttgtctg ctgtggacac tgagaaagt ggccatgagg 1260
cctgcttgcc cggggatcga gacagtagcc aagctccccg gcgagagccc caatgccgtc 1320
tgggggacgt ttgaggcggt ggcactagga gtgcacatct gtgagcatga caagcttatc 1380
ctcccatggt aacagaagtc caggctgagg ctgattctgg acgctgccct ttcagcacac 1440
gcagagcaaa gatcggttga agccccagtg tgggagatgc tcctcaggga ggaagccatg 1500
tgagggggct ggctctgtgg cgggtgagtg gtccccctct ccatcagcct ggacagccgc 1560
tcgggggtct aaggagtgc tcctgtcccg gcctggtgtg agtgggcagt gtaataaagt 1620
gtctttctat acggtgtcgc tcccacatc attttctcta gtgccgtgat tccttctaag 1680
aagactgact tccgtggccg ggcgcagtg ctcatgcctg taatcccagc actttgagag 1740
gccgaggttg ggagatcact tgaggtcagg agttcaagac cagcctggcc aacatgggtg 1800
aatcccatgt ctactaaaaa agacacaaat tagccaggcg tggtggcaca cacctgtagt 1860
cccagctacc tgggaggctg agacaggagg atcagctgaa cccgggaggt ggaggttgca 1920
gtgagccgag atcacaccac tgccctctag tattgtcact ggtgacaga gcgagactca 1980
gtctgaaaaa aaaaaa

```

&lt;210&gt; 250

&lt;211&gt; 1999

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 250

```

gggctagggc cggggcctgg ctgcgcggct gggccaaggc ccgcgatggt gatctgctgt 60
gcggccgtga actgtcccaa cggcaggga aaggcgaga agcgcgcgt ctccttccac 120
aggttcccc taaaggactc aaaacgtcta atccaatgt taaaagctgt tcagagggat 180
aactggactc cactaagta ttcatttctc tgtagtgagc atttcaccaa agacagcttc 240
tccaagaggc tggaggacca gcatcgctg ctgaagccca cggccgtgcc atccatcttc 300
cacctgaccg agaagaagag gggggctgga ggccatggcc gcacccggag aaaagatgcc 360
agcaaggcca cagggggtgt gaggggacac tcgagtgcgc ccacccggcag aggagctgca 420
ggttggtcac cgtcctcgag tggaaaccg atggccaagc cagagtccc caggttgaag 480
caagctgtc tgcaagggtg agccacaccc agggcgccc aggagggcgc cagccaggag 540
caggccagc aagctctgga acggactcca ggagatggac tggccaccat ggtggcaggc 600
agtcaggga aagcagaagc gtctgccaca gatgctggcg atgagagcgc cacttccctc 660
atcgaaaggg gcgtgacaga taagagtggc atttctatgg atgactttac gccccagga 720
tctggggcgt gcaaatttat cggctcactt cattcgtaca gtttctctc taagcacacc 780
cgagaaaagg catctgtccc ccgagagccc attgaccgca agaggctgaa gaaagatgtg 840
gaaccaagct gcagtggag cagcctggga cccgacaagg gcctggccca gaccctccc 900
agctcatcac ttaccgcgac accgcagaag ccttcccaga gcccctctgc cctcctgcc 960
gacgtcacc caaagccagc cacggaagcc gtgcagagcg agcacagcga cgccagcccc 1020
atgtccatca acgaggtcat cctgtcggcg tcaggggct gcaagctcat cgactcactg 1080
cactcctact gcttctctc ccggcagaac aagagccagg tgtgctgcct gcgggagcag 1140

```

```

gtggagaaga agaacggcga gctgaagagc ctgcggcaga gggtcagccg ctccgacagc 1200
caggtgcgga agctacagga gaagctggat gagctgagga gagtgagcgt cccctatcca 1260
agtagcctgc tgtcgcccag ccgagagccc cccaagatga acccagtggg ggagccactg 1320
tcctggatgc tgggcacctg gctgtcggac ccacctggag cggggaccta cccacactg 1380
cagcccttcc agtacctgga ggaggttcac atctcccacg tgggcccagcc catgctgaac 1440
ttctcgttca actccttcca cccggacacg cgcaagccga tgcacagaga gtgtggcttc 1500
attcgctca agcccgacac caacaagggtg gcctttgtca gcgcccagaa cacaggcgtg 1560
gtggaagtgg aggagggcga ggtgaacggg caggagctgt gcctcgcctc cactccatc 1620
gccaggatct ccttcgccaa ggagcccac gtagagcaga tcaccggaa gttcaggctg 1680
aattctgaag gcaaacctga gcagacggtc tccatggcaa ccacgacaca gccaatgact 1740
cagcatcttc acgtcaccta caagaagggt accccgtaaa cctagagctt ctggagccct 1800
cgggagggcc tggctactgt gcctcaacgg ttcggctcct caacagacag tccctgcggc 1860
aaaagtgggt gtggcctgta gcctctgcag gctcaagagt gttgtccaga tgtttctgta 1920
ctggcataga aaaaccaaat aaaaggcctt tatttttatg gctgaggatt ttgaatatta 1980
aaaaaaaaa aaaaaaaaaa

```

&lt;210&gt; 251

&lt;211&gt; 1398

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 251

```

ggctgtgcgc cacttccggc ttcaaccccc gaaaaggcgg tgcttaaacc ggaggaggcg 60
gaagtgagtc gacagacgag gcggctttcc cggcagaatg ctagcgcagg cgcaggggct 120
cgagaggcct ggacctgtgg cgcctcctca gtgaggaggg ccgcccgtga tccgtcgccg 180
gccccggtct ccaggggcct caccgagtc atgccccgt attgcgcagc gatttgtgt 240
aagaaccgcc ggggacgaaa caataaagac cggaagctga gtttttatcc atttctcta 300
catgacaaag aaagactgga aaagtgggta aagaatatga agcgagattc atgggttccc 360
agtaaatacc agtttctatg tagtgaccat tttactcctg actctcttga catcagatgg 420
ggtattcgat atttaaaaca aactgcagtt ccaacaatat tttctttgcc tgaagacaat 480
cagggaaaag acccttctaa aaaaaaatcc cagaagaaaa acttgaaga tgagaaagaa 540
gtatgcccaa aagccaagtc agaagaatca tttgtattaa atgagacaaa gaaaaatata 600
gttaacacag atgtgcccc tcaacatcca gaattacttc attcatcttc ctgggtaaa 660
ccaccagctc caaaaacagg aagtatacaa aataacatgt taactcttaa tctagttaaa 720
caacatactg ggaaaccaga atctaccttg gaaacatcag ttaaccaaga tacaggtaga 780
ggtgggtttt acacatgttt tgagaatcta aattctacaa ctattacttt gacaacttca 840
aattcagaaa gtattcatca atctttggaa actcaagaag ttcttgaagt aactaccagt 900
catcttgcta atccaaactt tacaagtaat tccatggaaa taaagtcagc acaggaaaa 960
ccattcttat tcagcacaat taatcaaaca gttgaagaat taaacacaaa taaagaatct 1020
gttattgcca tttttgtacc tgctgaaaat tctaaaccct cagttaattc ttttatatct 1080
gcacaaaaag aaaccacgga aatggaagac acagacattg aagactcctt gtataaggat 1140
gtagactatg ggacagaagt tttaaaaatc gaacattctt actgcagaca agatataaat 1200
aaggaaatc tttggcagaa agtctctaag ctacattcaa agataactct tctagagtta 1260
aaagagcaac aaactctagg tagattgaag tctttggaag ctcttataag gcagctaaag 1320
caggaaaact ggctatctga agaaaacgct aagattatag aaaaccattt tacaacatat 1380
gaagtcacta tgatatag

```

&lt;210&gt; 252

&lt;211&gt; 2291

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 252

```

agcgaaggca gacgcagtct ccctcgttga cgtagtgcgc agtcttcgct gctaacggtt 60
tgttatgagt tgctaaaatg gtgaaatgct gctccgccat tggatgtgct tctcgtgct 120
tgccaaattc gaagttaaaa ggaactgacat ttcacgtatt cccacagat gaaaacatca 180
aaaggaaatg ggtattagca atgaaaagac ttgatgtgaa tgcagccggc atttgggagc 240
ctaaaaaagg agatgtgttg tgttcgaggc actttaagaa gacagatttt gacagaagt 300
ctccaaatat taaactgaaa cctggagtc taccctctat ctttgattct ccatatcacc 360
tacaggggaa aagagaaaaa cttcattgta gaaaaaactt caccctcaa accgttccag 420
ccactaacta caatcccat cttgttggtg cttcctcatg tattgaagaa ttccaatccc 480

```

```

agttcatttt tgaacatagc tacagtgtaa tggacagtcc aaagaaactt aagcataaat 540
tagatcatgt gatcggcgag ctataggata caaaggaaag tctacggaat gtttttagacc 600
gagaaaaacg ttttcagaaa tcattgagga agacaatcag ggaattaaaag gatgaatgtc 660
tgatcagcca agaaacagca aatagactgg acactttctg ttgggactgt tgtcaggaga 720
gcatagaaca ggactatatt tcatgaaata atttcatgtt acgttccacc taaaattgtc 780
attgggtacaa atttttataa aatctcattt accatcacta aataatatcc atcattttaa 840
gtgctgcttt ggatttctctg gagcattatg cattatagtt gttatccaaa gacttttttg 900
aaaaatatga gaaattttgt gtaattatgt atttgtgtct tgtgacaatt atgttttata 960
gacctacact agtgccaggt cactattgta agatgttaaa atctcaagaa aatttcacag 1020
agctaaagaa atgatgtcaa attagtcaca ttaagctata gtagaaggaa ttggacactt 1080
ctccagatat ttggcttcaa aggagtacct ttacttacat gtgctttatg gtaagtacat 1140
tgaattttac tttaaatgca ttttactaca aagcacaatt cttttgtaat gcatatccat 1200
cttggattca atccaagggtg ctttagctat cagtagtacc aaaggatctt tttacaaggc 1260
ttcctgtggt attgactctg agaataacac atagtgaaga tctgtgggct tttaaaattg 1320
ttcacagcca atttaagaag acccctcatg aagtctcagt tttcagtaca gtacatcatt 1380
cctcctcact aggagcactt tgatgtaaag cagaatagct ttaaaaagac aaaaaggatc 1440
gtagatctga tttttaaatg gttggttgct ctgacagatc tgaacacttt gcttcattac 1500
tatttcgtca taaaggata tgtttaaaat ctgaatggca gtactagctc tatactttta 1560
atactgcttt gtattttata tgtaaagtag tattgctgac attttaaaaa aatacaaaat 1620
acaaaagaaa ccattagaaa ttaataactg tggctcttcc agttgaaata ggaattggag 1680
agaaaggatt agaataattt aattagggga gtagattatt gtccaaaggc ttttatttag 1740
agaaacgggt aattaaaaca gcagctttag aatagcttct tactgaatat gcaaaagaat 1800
aattccttgt tatttcctaa ttgatccaa tctcataaat ttagcttttg tcataattcc 1860
ttaccgaaaa caactgaaat tgagagtcac aaatactgtg ggtagaata aaaaccattt 1920
gccaaagcaa cactctactt agaagcacat gtacatacat ggacctcatt cagaagtcca 1980
tggtgtagca gttagaattt gagtatcagc catttcattg tagtaacaaa aattgaattg 2040
cattttgtgc tcagtgtttt attgtaattt tatttttgtt acattaatat tagttaagat 2100
atggtcactt gaattttttg tatttaagaa ttttctgttt taatgcatgt tatactttta 2160
tgtaggattc ccaaccttcc ctctaaatgg gatttaaccc acatctgcga gatcagcgtt 2220
atgctaagag gaatcactg aggccatata tttttacaat ctgaaaaaaa agtagtaaaa 2280
aggtagttaa a 2291

```

&lt;210&gt; 253

&lt;211&gt; 1242

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 253

```

cgtgagtgcc gctgacagaa gtcaagagaa tccggtggga cggggttggg gcgacaacgg 60
gccggggggg acccgacagg ccagagcccc ttggggagga gcggcggtg gaggcgcgag 120
gctcctccgg atgcccggag agcgcgttgc gacttaactc ccgcctctt cccagatgcc 180
gcgtcactgc tccgcgcgcg cctgctgcac acgggacacg cgcgagacgc gcaaccgcgg 240
catctccttc cacagacttc ccaagaagga caaccgagg cgaggcttg ggctggccaa 300
ctgccagcgg ctggacccca gcggccaggg cctgtgggac ccggcatccg agtacatcta 360
cttctgtctc aaacactttg aggaggactg ctttgagctg gtgggaatca gtggatatca 420
caggctaaag gagggggcag tccccaccat atttgagtct ttctccaagt tgcgccggac 480
aaccaagacc aaaggacaca gttaccacc tggccccctt gaagtcagcc ggctcagacg 540
atgcaggaag cgctgctccg agggccgagg gccacaact ccattttctc cacctccacc 600
tgctgatgtc acctgctttc ctgtggaaga ggctcagca cctgccactt tgccggcctc 660
cccagctggg aggtctggag ctggccttag cagccccttt tcagacctac tgggcccctt 720
gggtgcccag gcagatgaag caggctgcag cgcccagcct tcaccagagc ggcagccctc 780
ccctctcgaa ccacggccag tctccccctc agcgtatatg ctgcgcctgc ccccaccgcg 840
cggagcctac atccagaatg aacacagcta ccagggtggc agcgccttac tctggaagcg 900
gcgagccgag gcagcccttg atgcccttga caaggcccag cgccagctgc aggcctgcaa 960
gcggcgggag cagcggctgc ggttgagact gaccaagctg cagcaggagc gggcacggga 1020
gaagcgggca caggcagatg cccgccagac tctgaaggag catgtgcagg actttgccat 1080
gcagctgagc agcagctgg cctgaggggc tgctggactg accgaggggc tgcccagcaa 1140
gactgcagcc tcttctctcc tcagatccca ccagaccac cagggtgcat aataaagcgg 1200
attctagacg gagaaaaaaa aaaaaaaaaa aaaaaaaaaa aa 1242

```

&lt;210&gt; 254

<211> 1383  
 <212> DNA  
 <213> Homo sapiens

<400> 254

```

agcgggggtc ggggttaggc ggcgctccgc gagaacccaa gtgcagccgc tgaccgggca 60
aaactcagcg ggggctggat agccatgccc aagtactgca gggcgccgaa ctgctccaac 120
actgcggggc gcctgggtgc agacaaccgc cctgtgagct tctacaagtt cccactgaag 180
gatggtcccc ggctgcaggc ctggctgcag cacatgggct gtgagcactg ggtgcccagc 240
tgccaccagc acttgtgcag cgagcacttc acaccctcct gcttccagtg gcgctggggg 300
gtgcgctacc tgcggcctga tgcagtgcgc tccatcttct cccgggggacc acctgccaa 360
agtcagcgga ggacccgaag caccagaag ccagtctcgc cgccgcctcc cctacagaag 420
aatacaccct tgcgccagag ccctgccatc ccagtctctg gccagtgcg cctagtggg 480
ctgggccccca catcggggag ccccaagact gtggccacca tgctcctgac cccctggcc 540
cctgcgccaa ctctgagcg gtcacaacct gaagtccctg cccaacaggc ccagaccggg 600
ctgggccccag tgctgggagc actgcaacgc cgggtgcgga ggctgcaacg gtgccaggag 660
cggcaccagg cgcagctgca ggccctggaa cggtggcac agcagctaca cggggagagc 720
ctgctggcac gggcagcgc gggctctgag cgctgaca cagcccagac ccttgaccc 780
gaggaatccc catctctcac catcatctgt ggaggcctg acatagccat ggtccttgcc 840
caggaccctg cacctgccac agtggatgcc agggacaaa gatagaagat ggaggaggaa 960
agtgcataag gatcaagaca gacaatgtcg agggacaaa gatagaagat ggaggaggaa 960
agacattata cgtgggcttg gccagcccc accgcccacg cctgggtagt agcagtgcc 1020
ccctcaaggg cctgggttct accaccccac tcctagggat ctcttgaaac ttaggggtga 1080
cctgggcccc agtctctcat cagcccccaa tccccgggt accaggcttc tgccacccc 1140
ggctcagatc ttgcaaatc agtacgacag cctcagagca gagcaagggt tgtttgggag 1200
aatcatacct ggttctaagg agtcccacg tttttgcaa gcctggtagt gagtctatga 1260
taccatggtg gacacagctg agaaaatccc tgccctcatg gtgctcattc tacttgagta 1320
gacgatgaac tagtaacaa ataaacaaga aactgcaga catgaaaaaa aaaaaaaaaa 1380
aaa
1383

```

<210> 255  
 <211> 3627  
 <212> DNA  
 <213> Homo sapiens

<400> 255

```

attcatgctg tcgcggaac cccgaagggtg gggccccacg taacaagaag atgaccggaa 60
gttgctccgc agtgggctgc agcaccgtg acaccgtgct cagccgggag cgcggcctct 120
ccttccacca atttccaact gataccatac agcgtcmeta atggatcagg gctgttaatc 180
gtgtggaccc cagaagcaaa aagatttggg ttcaggacc aggtgctata ctgtgttcca 240
aacattttca agaaagtgc tttgagtcac atggcataag aagaaagctg aaaaaaggag 300
ctgtgccttc tgtttctcta tacaagattc ctcaagggtg acatcttaaa ggtaaagcaa 360
gacaaaaaat ctaaaaacaa cctcttccag acaattctca agaagttgct actgaggacc 420
ataactatag tttaaagaca ctttgacga taggtgcaga gaaactggct gaggtgcaac 480
aatgtttaca agtgtccaaa aaaagactta tctccgtaaa gaactacagg atgatcaaga 540
agagaaaggg tttacgatta attgatgcac ttgtagaaga gaaactactt tctgaagaaa 600
cagagtgtct gctacgagct caattttcag attttaagtg ggagttatat aattggagag 660
aaacagatga gtactccgca gaaatgaaac aatttgcag tacactctac ttgtgcagta 720
gcaaagtcta tgattatgta agaaagattc ttaagctgcc tcattcttcc atcctcagaa 780
cgtggttatc caaatgccaa ccagctccag gtttcaacag caacattttt tcttttcttc 840
aacgaagagt agagaatgga gatcagctct atcaatactg ttcattgtta ataaaaagta 900
tacctctcaa gcaacagctt cagtgggac ctagcagtc cagtttgag gggtttatgg 960
actttggctc tggaaaactt gatgctgatg aaacgccact tgcttcagaa actgttttgt 1020
taatggcagt gggtattttt ggccattgga gaacacctct tgggtatttt tttgtaaaaca 1080
gagcatctgg atatttgag gctcagctgc ttcgtctgac tattggtaaa ctgagtgaca 1140
taggaatcac agttctggct gttacatctg atgccacagc acatagtggt cagatggcaa 1200
aagcattggg gatgagacat gatggagac acatgaaatg tacatttcag catccttcac 1260
cttctagtca acagattgca tacttctttg actcttgcca cttgctaaga ttaataagaa 1320
atgcatttca gaattttcaa agcattcagt ttattaatgg tatagcacat tggcagcacc 1380
tcgtggagtt agtagcactg gaggaacagg aattatcaaa tatggaaaga ataccaagta 1440
cacttgcaaa tttgaaaaat catgtactga aagtgaatag tgccacccaa ctctttagtg 1500

```



```

agagtgtagc cagtgcatta gaatatattgt tacccttaga cctgccacct tttcaaaact 1560
gtattggtac catccatttt ttacgttttaa ttaacaatct gtttgacatc tttaatagta 1620
ggaactgtta tggaaaggga cttaaagggc ctctgttgcc tgaaacttac agtaaaataa 1680
accacgtgtt aattgaagcc aagactatct ttgttacatt atctgacact agcaataatc 1740
aaataattaa aggtaagcaa aaactaggat tcttgggatt tttgctcaat gctgagagct 1800
taaaatggct ctaccaaatt tatgttttcc caaaggctcat gccttttcct tatcttctga 1860
cttacaattt cagtcattgat catctggaat tatttctaaa gatgcttagg caggatttag 1920
taacaagttc tagccctacc tgcattggat tccagaaagc ttactataat ttggagacca 1980
gatacaaat tcaagatgaa gtttttctaa gcaaagtaag catctttgac atttcaattg 2040
ctcgaaggaa agacttggcg ctttggacag ttcaacgtca gtatggtgtc agcgttacaa 2100
agactgtctt tcacgaagag ggtatttgtc aagactggtc tcattgttca ctaagtgagg 2160
cattactaga cctgtcagat cataggcgaa atctcatctg ttatgctggt tatgttgcaa 2220
acaagttatc agctctttta acttgtgagg actgcatcac tgcactgtat gcatcggatc 2280
tcaaagcctc taaaattggg tcaactattt ttgttaaaaa gaagaatggt ttgcattttc 2340
cttcagaaag tctgtgtcgg gtcataaata ttgttgagcg agttgtaaga acccattcaa 2400
gaatggcaat ttttgaaact gtttctaaac aaagggaatt gtatcttcaa cagaaaatat 2460
tatgtgagct ttctgggcat attgatcttt ttgtagatgt gaataagcat ctctttgatg 2520
gagaagtgtg tgccatcaat cactttgtca agttgctaaa ggatataata atctgtttct 2580
taaataatcag agctaaaaat gttgcacaga atcctttaaa acatcattca gagagaactg 2640
atatgaaaac tttatcaagg aaacactggg cacctgtaca ggattataaa tgttcaagtt 2700
ttgctaatac cagtagtaaa ttccaggcat tgctaagtaa cgatggatat ccattcaaat 2760
gagagacctt aaatatatta acattttaat taagaatact tgatcaacat tttttgaagt 2820
tcaattttacc atattttata aattggcgat tctgcacagt ggacaagttt gcaattctga 2880
cttattaaaa tttcaaatc tgcataatcac aaaatctcct tatacttttg gtatggcttg 2940
cagcatttat gagttttcca aaatatagaa agcagtaggt cagtaggagc aaactagcca 3000
acaggtagct tctttgaatt tactactgta agactaagca gtgttactgg acacagtttt 3060
aacttgttca atctgtttca aaaacaagaa aaacaacaac tatgagttat caaaatattg 3120
actccattta tgactagact acatttctga aagatctttg gtttacgatt cttaagaata 3180
ttgacaatac ctataaaact ttgaagataa cttttactta aatatgaaaa ttatagtttg 3240
aaaattaggg tcaagcaaat atcaaatact gcaaaaatcc ccttgtccca ggatacccta 3300
aaatagaagt atattttgat gtttgtttat tctacctcaa acagaggctt aagttttgaa 3360
gtgtaaccag tttatacttc attttataca aaacttattg ctgagaagtc tgaatattgt 3420
gctttttgtt gtttgtaaat agaattgaat ttaaatcac caggataaat cttattttaa 3480
ggaagcctgt ttgaaaatca ccaactttaa cttattgctt atataaatcc aagctctgta 3540
cctgatcttt atgtaaagca agattcatat gtgtagtatc taatgccctt tgggtgttaca 3600
tttgactaaa atacaaatgt ctgtattt 3627

```

&lt;210&gt; 256

&lt;211&gt; 771

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 256

```

atgccggccc gttgtgtggc cgccactgac ggcaacacca ccaagtctgg gaagtcgctg 60
ttccgctttc ccaaggaccg ggccgtgagg ctgctctggg accgcttctg gcgggggttg 120
cgcccgact ggtacggagg caatgaccgc tcggtcatct gctctgacca ctttgcccca 180
gcctgttttg acgtctcttc gggtatccag aagaacctgc gcttctccca gcgcctgagg 240
ctggtggcag gcgcctgacc caccctgcac cgggtgcccg ccccggcacc taagagggga 300
gaggagggag accaagcagg ccgcctggag acgcgaggag agctccaggc agccaggcat 360
tctgaggctg cccaggtcc agtctcctgt acacgcccc gagctgggaa gcaggctgca 420
gcttcacaga ttacgtgtga aaatgaactt gtgcaaacc aaccccatgc tgataatcca 480
tctaatactg tcaattcagt acctactcac tgtgaagaag gccagtgca taaaagtaca 540
caaatttctt tgaaaaggcc ccgtcaccgt agtgtgggta tcaagccaa agtgaaagcg 600
tttgaaaaaa gactgtgtaa tgcaactact cagacagagg aattgtggtc tagaacttcc 660
tctctctttg acatttactc cagtgattca gaaacagata cagactggga tatcaagagt 720
gaacagagtg atttgtctta tatggctgta caggtgaaag aagaaacatg t 771

```

&lt;210&gt; 257

&lt;211&gt; 942

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 257

```

atgcttggtt ttacgtgctg cgtgccaggc tgctacaaca actcgcaccg ggacaaggcg 60
ctgcacttct acacgtttcc aaaggacgct gagttgcggc gcctctggct caagaacgtg 120
tcgctgcccg gcgtcagtggt gtgctttctc accttccagc ccaccacagg ccaccgtctc 180
tgcagcgttc acttccaggc cggccgcaag acctacacgg tacgcgtccc caccatcttc 240
ccgctgcgcg gcgtcaatga gcgcaaagta gcgcgcagac ccgctggggc cgcggccgcc 300
cgccgcaggc agcagcagca acagcagcag cagcagcaac agcagcaaca gcagcagcag 360
cagcaacagc agcagcagca gcagcagcag cagcagtcct caccctctgc ctccactgcc 420
cagactgccc agctgcagcc gaacctggta tctgcttccg cggccgtgct tctcaccctt 480
caggccactg tagacagcag tcaggctccg ggatccgtac agccggcgcc catcactccc 540
actggagaag acgtgaagcc catcgatctc acagtgaag tggagtttgc agccgcagag 600
ggcgcagccg ctgcccggcg cgcgtcggag ttacaggctg ctaccgcagg gctggaggct 660
gccgagtgcc ctatgggccc ccagttggtg gtggtagggg aagagggtt cctgataact 720
ggctccgacc attcgtactc cttgtcgtca ggcaccacgg aggaggagct cctgcgcaag 780
ctgaatgagc agcgggacat cctggctctg atggaagtga agatgaaaga gatgaaaggc 840
agcattcgcc acctgcgtct cactgaggcc aagctgcgcg aagaactgcg tgagaaggat 900
cggctgcttg ccatggctgt catccgcaag aagcacgga tg 942

```

&lt;210&gt; 258

&lt;211&gt; 2283

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 258

```

atgccgaact tctgcgtgc ccccaactgc acgcggaaga gcacgcagtc cgacttgccc 60
ttcttcaggt tcccgcgga cctgccaga tgccagaagt gggtaggaga ctgtaggaga 120
gcagacttag aagataaaac acctgatcag ctataaaac attatcgatt atgtgcaaaa 180
cattttgaga cctctatgat ctgtagaact agtccctata ggacagttct tcgagataat 240
gcaataccaa caatatgtga tcttaccagt catttgaaca acccacatag tagacacaga 300
aaacgaataa aagaactgag tgaagatgaa atcaggacac tgaaacagaa aaaaattgat 360
gaaacttctg agcaggaaca aaaacataaa gaaaccaaca atagcaatgc tcagaacccc 420
agcgaagaag aggggtgaagg gcaagatgag gacattttac ctctaaccct tgaagagaag 480
gaaaacaaag aatacctaaa atctctatct gaaatcttga ttctgatggg aaagcaaaaac 540
atacctctgg atggacatga ggctgatgaa atcccagaag gtctctttac tccagataac 600
tttcaggcac tgctggagtg tcggataaat tctgggtgaag aggttctgag aaagcggttt 660
gagacaacag cagttaacac gttgttttgt tcaaaaacac agcagaggca gatgctagag 720
atctgtgaga gctgtattcg agaagaaact ctacgggaag tgagagactc acacttcttt 780
tccattatca ctgacgatgt agtggacata gcaggggaag agcacctacc tgtgttggtg 840
aggtttgttg atgaatctca taacctaaag gaggaattta taggcttctc gccttatgaa 900
gccgatgcag aaattttggc tgtgaaattt cacactatga taactgagaa gtggggatta 960
aatatggagt attgtcgtgg ccaggcttac attgtctcta gtggattttc ttccaaaatg 1020
aaagtgttg cttctagact tttagagaaa tatccccaag ctatctacac actctgctct 1080
tctgtgcct taaatatgtg gttggcaaaa tcagtaacctg ttatgggagt atctgttgca 1140
ttaggaacaa ttgaggaagt ttgttctttt ttocactgat caccacaact gcttttagaa 1200
cttgacaacg taatttctgt tctttttcag aacagtaaaag aaaggggtaa agaactgaag 1260
gaaatctgcc attctcagtg gacaggcagg catgatgctt ttgaaatttt agtggaactc 1320
ctgcaagcac ttgttttatg ttttagatgg ataaatagtg acacaaatat tagatggaat 1380
aactatatag ctggccgagc atttgtactc tgcagtgcag tgtcagattt tgatttcatt 1440
gttactattg ttgttcttaa aaatgtccta tcttttacia gagccttttg gaaaaacctc 1500
caggggcaaa cctctgatgt cttcttttgc gccggtagct tgactgcagt actgcattca 1560
ctcaacgaag tgatggaaaa tattgaagtt tatcatgaat tttgttttga ggaagccaca 1620
aatttgga ccaaaactga tattcaaatg aaactccctg ggaaattccg cagagctcac 1680
cagggtaact tggaatctca gctaacctct gagagttact ataaagaaac cctaagtgtc 1740
ccaacagtgg agcacattat tcaggaactt aaagatatat tctcagaaca gcacctcaaa 1800
gctcttaaat gcttatctct ggtacctca gtcattgggac aactcaaat caatacgtcg 1860
gaggaacacc atgctgacat gtatagaagt gacttaccca atcctgacac gctgtcagct 1920
gagcttcatt gttggagaat caaatggaaa caccggggga aagatataga gcttccgtcc 1980
accatctatg aagccctcca cctgctgac atcaagtttt ttcctaattg gtatgcattg 2040
ctgaaggctc tgtgtattct tctgtgatg aaggttgaga atgagcggtg tgaaaatgga 2100
cgaaagcgtc ttaaagcata tttgaggaac actttgacag accaaaggte aagtaacttg 2160

```

gctttgctta acataaatTT tGatataaaa cagcactggt atttaattggt ggacacatat 2220  
 attaaactct atacaagtaa gtcagagctt cctacagata attccgaaac tgtggaaaaat 2280  
 acc 2283

<210> 259  
 <211> 986  
 <212> DNA  
 <213> Mus musculus

<400> 259  
 cttctgctaa agcaaaccac acaacggaca gggtagtcac tcgcccaccc caacccccac 60  
 cccacggcga ggtgatcgct cccgtaactg ctgaccgacg ccaccgagag cggcgagcgt 120  
 tatcaaggcc gagcgcgga ccccgacggc ccccttcgcc tgccctcccg gccgaaggag 180  
 agtgtggagg gccagaagga tgggtgcagtc ctgctccgcc tacggtgca agaaccgcta 240  
 cgacaaggac aagcccgtct ccttcacaa gtttctctt actcgcccca gcctttgtaa 300  
 gcagtgggag gcagctgtta aaaggaaaaa cttcaagccc accaagtaca gcagcatctg 360  
 ctcgagcac ttcaccccg actgctttaa gagggagtgc aacaacaagc tactgaagga 420  
 gaacgtgtg cccacaatat ttctctatat cgagccacat gagaagaagg aagacctgga 480  
 atcccaagaa cagctccctt ctctctcacc ccccgcttc caggttgatg ctgctattgg 540  
 gctgctaattg cccctctgc agacccctga taacctgtcg gttttctgtg accacaatta 600  
 cactgtggag gatacgatgc accagaggaa gaggatcctg cagctggagc agcaggtgga 660  
 gaaactcagg aagaagctca agacggccca gcagcggtgc cggcggcagg agaggcagct 720  
 cgagaagctc aaggaagctg tccactttca gagagaagag gacgacgctg ccgagagggg 780  
 ctacgtgatc ctaccaaag actactttga aattgttgaa gttccagcat gaaaaaatga 840  
 gatgtgttag tgggacaaga ctatacacct tcttttagcc tacatacagg agttcatttg 900  
 aaaaaataac acttaattac ttgtattaaa aaaacaatat ttttttaaaa taaattagat 960  
 atatactgta aaaaaaaaa aaaaaa 986

<210> 260  
 <211> 1515  
 <212> DNA  
 <213> Mus musculus

<400> 260  
 gctctgccct ccccgcgctc tgcaccgagc tggcgggcgc gggtcgctg cctcggttgt 60  
 ctacggtttg acagaagctt gcttagcggg cagcgccctc gaagtggcgt aaggtggcgc 120  
 cgaatgaggg gcccggggaa atgccgacca attgcgcgcg ggcgggctgt gctgctacct 180  
 acaacaagca cattaacatc agcttccaca ggtttccttt ggatcctaaa agaagaaaag 240  
 aatgggttcg cctggttagg cgcaaaaatt ttgtgccagg aaaacacact tttctttgct 300  
 caaagcactt tgaagcctcc tgttttgatc taacaggaca aaccgacga cttaaaatgg 360  
 atgctgttcc aaccattttt gatttttgta cccatataaa gtctctgaaa ctcaagtcaa 420  
 ggaatcttct gaagacaaac aacagttttc ctccaactgg accatgtaat ttaagctga 480  
 acggcagtc gcaagtactg cttgaacaca gttatgcct taggaaccct atggaggcga 540  
 aaaaaaggat aattaaacta gaaaaggaaa tagcaagctt gagaaaaaaa atgaaaactt 600  
 gcctgcaaag agaacgcaga gcaactcgaa ggtggatcaa agccacgtgc tttgtgaaga 660  
 gcttagaagc aagtaacatg ctacctaaag gcatctcaga acagatttta ccaactgcct 720  
 taagcaatct tcctctggaa gatttaaaaa gtcttgaaca agatcaacaa gataaaacag 780  
 taccatttct ctaaatgtaa aatggaagag actctctgca ctcaagtttt cctcacacag 840  
 aaccagtgct cagctcctg ccgtcccac ccaccgact ctgacagtta cactacaatc 900  
 aagtcctgca gttttacttg aagtagtagt gtcagtgtca ctctctggag actgagggaag 960  
 tgggaaatcc aatgacaagc ttgacaccga gcagaagtgc cttacatgag ggtcacggac 1020  
 ttaggaaaca ctgccagcag ggtttctgct cttgtttttt taagctgctg tcaaatagga 1080  
 atgacaagtg atatgttcat aaaagtaaaa gcattccgca ccaaagctgg gatattacat 1140  
 tctaaagaac atgtgaagta ggagctaact gcattaaata tgatcttaaa actactaatg 1200  
 tatttttgtat gaattaaatt attgggattg tgggtgaaaa ttttatagaa taaaacctct 1260  
 ggggtacggg gcaaggtttg ttctttgtt ttgtttgtt tttgtcttt tttagccttt 1320  
 tgtattttaa ctagtaaaag taaacttatt atggcctttt tttataagaa cattgaattt 1380  
 aaaagtaggt tgtaaaataa tctgaaatag tattttgaat gtgaaatacc tttgaaactc 1440  
 caaactaggt aaggcccaa gcacctcaga ctgggaaaac ccagtgaagt atagtcaacg 1500  
 tctaagaaaa tatac 1515

<210> 261  
 <211> 1120  
 <212> DNA  
 <213> Mus musculus

<400> 261  
 gaggggacagt gggcccatct ccgagatgcc gaagtcttgc gcggcccgcc aatgctgcaa 60  
 ccgctacagc agccgcagga agcagctcac cttccaccgg tttcccttca gccgcccaga 120  
 gctgttgagg gagtgggtgc tcaacatcgg ccgggctgac ttcaagccta agcagcacac 180  
 agtcatctgc tcggaacact tcagaccgga gtgcttcagc gcctttggga accgcaagaa 240  
 cctgaaacac aatgctgtgc ccacggtgtt cgcttttcag aacccacacag aggtctgccc 300  
 tgaggtgggg gctggtgggg acagctcagg gaggaacatg gacaccacac tggaagaact 360  
 tcagcctcca accccggaag gccccgtgca gcaggtctta ccagatcgag aagcaatgga 420  
 ggccacggag gccgctggcc tgcctgccag ccctctgggg ttgaagaggc cccttcgggg 480  
 acagccgtct gatcacagtt atgccctttc ggacttggat accctcaaaa aaaaactctt 540  
 tctcacactg aaggaaaaca agaggcttcg gaagcggctg aaagcccaga ggctgctgtt 600  
 gcggaggaca tgtggccgcc tgagagccta cagagaggga cagccgggac ctccggccag 660  
 acggccggca cagggaagct gagcctgagc aagctctggg atgtgggggt ggtggcaaca 720  
 ccttagcagg aagtgggtgt ctggcctgct atgggcgttt ctaccgctg ctgatgctgc 780  
 aggtgccttg agagtgggat gggatgctgc gacaggcagt tgcgggtgg gggcccaagt 840  
 actgcccagg caccgtccca ggtttcttgg gctgaggctg tcagctgtgg ggaagcagca 900  
 gtgaccaa atgtgagccgtc acaaccccct caagagatgc tcccagaggg agagctggtc 960  
 attccttacag ccggtgggggt ccttactgtc tccccatagg agccattctg atggcaggca 1020  
 gggcaagggt ccccgtcagc ctgtatttct gagtgaactt tttttctgcc tggttcgtgt 1080  
 agatgtggaa taaatctttt gaagtctcca aaaaaaaaaa 1120

<210> 262  
 <211> 558  
 <212> DNA  
 <213> Mus musculus

<400> 262  
 atactgcaag catttggaag cctaaaaaaaa ggagatgtgc tgtgttcaag acacttcaag 60  
 aagacagact ttgacagaag cactctaaac actaagctga aggcaggagc catcccttct 120  
 atctttgaat gtccatatca cttacaggag aaaagagaaa aacttcactg tagaaaaaac 180  
 ttcctttctca aaacccttcc catcacccac catggccgcc agcttgttgg tgcctcctgc 240  
 attgaagaat tcgaacccca gttcattttt gaacatagct acagtgttat ggacagccca 300  
 aagaagctta agcataagct agaccgtgtg atcatcgagc tggagaatac caaggaaagc 360  
 ctacggaatg ttttagcccg agaaaaacac tttcaaaagt cactgaggaa gacaatcatg 420  
 gaactaaagg atgaaagtct gatcagccag gaaacagcca atagtctggg tgctttctgt 480  
 tgggagtgtc atcatgaaag cacagcagga ggctgtagtt gtgaagtcac ttcttatatg 540  
 cttcatctgc agttgaca 558

<210> 263  
 <211> 37  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Consensus sequence for PAR4 binding domain of THAP

<221> UNSURE  
 <222> 3-15, 18, 22, 24-34, 36  
 <223> Xaa = any of the twenty amino acids

<400> 263  
 Leu Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Gln  
 1 5 10 15  
 Arg Xaa Arg Arg Gln Xaa Arg Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
 20 25 30  
 Xaa Xaa Gln Xaa Glu

35

<210> 264  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 264  
ccgctcgagg tgcagtcctg ct 22

<210> 265  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 265  
cgggatccgc tggctattca actatttca 29

<210> 266  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 266  
ccgctcgagg atacaatgca cc 22

<210> 267  
<211> 33  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 267  
gcgggatccg ctggtacttc aactatttca aag 33

<210> 268  
<211> 86  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Synthetic Oligonucleotide

<400> 268  
ccgctcgagc caccatggag acagacacac tctgctatg ggtactgctg ctctgggttc 60  
caggttccac tggtgacctc gagatt 86

<210> 269  
<211> 26

<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 269  
taggggtcgac gccacatgg agacag

26

<210> 270  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 270  
ccgctcgagg tcaccagtgg a

21

<210> 271  
<211> 134  
<212> PRT  
<213> Homo sapiens

<400> 271  
Met Ala Gln Ser Leu Ala Leu Ser Leu Leu Ile Leu Val Leu Ala Phe  
1 5 10 15  
Gly Ile Pro Arg Thr Gln Gly Ser Asp Gly Gly Ala Gln Asp Cys Cys  
20 25 30  
Leu Lys Tyr Ser Gln Arg Lys Ile Pro Ala Lys Val Val Arg Ser Tyr  
35 40 45  
Arg Lys Gln Glu Pro Ser Leu Gly Cys Ser Ile Pro Ala Ile Leu Phe  
50 55 60  
Leu Pro Arg Lys Arg Ser Gln Ala Glu Leu Cys Ala Asp Pro Lys Glu  
65 70 75 80  
Leu Trp Val Gln Gln Leu Met Gln His Leu Asp Lys Thr Pro Ser Pro  
85 90 95  
Gln Lys Pro Ala Gln Gly Cys Arg Lys Asp Arg Gly Ala Ser Lys Thr  
100 105 110  
Gly Lys Lys Gly Lys Gly Ser Lys Gly Cys Lys Arg Thr Glu Arg Ser  
115 120 125  
Gln Thr Pro Lys Gly Pro  
130

<210> 272  
<211> 878  
<212> DNA  
<213> Homo sapiens

<400> 272  
atcccagccc acgcacagac ccccaacttg cagctgccca cctcaccctc agctctggcc 60  
tcttactcac cctctaccac agacatggct cagtcaactg ctctgagcct ccttatcctg 120  
gttctggcct ttggcatccc caggacccaa ggcagtgatg gaggggctca ggactgttgc 180  
ctcaagtaca gccaaaggaa gattccgcc aaggttgctc gcagctaccg gaagcaggaa 240  
ccaagcttag gctgctccat ccagctatc ctgttcttgc cccgcaagcg ctctcaggca 300  
gagctatgtg cagacccaaa ggagctcttg gtgcagcagc tgatgcagca tctggacaag 360  
acaccatccc cacagaaacc agcccagggc tgcaggaagg acaggggggc ctccaagact 420  
ggcaagaaag gaaagggtc caaagggtgc aagaggactg agcggtcaca gaccctaaa 480  
gggcatagc ccagtgcagc gcttgagacc ctggagaccc caccagcctc accagcgctt 540

```

gaagcctgaa cccaagatgc aagaaggagg ctatgctcag gggccctgga gcagccaccc 600
catgctggcc ttgccacact ctttctcctg ctttaaccac cccatctgca ttcccagctc 660
taccctgcat ggctgagctg cccacagcag gccagggtcca gagagaccga ggaggggagag 720
tctcccaggg agcatgagag gaggcagcag gactgtcccc ttgaaggaga atcatcagga 780
ccctggacct gatacggctc cccagtacac cccacctctt ccttgtaaata atgatttata 840
cctaactgaa taaaaagctg ttctgtcttc ccacccaa 878

```

&lt;210&gt; 273

&lt;211&gt; 98

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 273

```

Met Ala Leu Leu Ala Leu Ser Leu Leu Val Leu Trp Thr Ser Pro
 1           5           10          15
Ala Pro Thr Leu Ser Gly Thr Asn Asp Ala Glu Asp Cys Cys Leu Ser
          20           25           30
Val Thr Gln Lys Pro Ile Pro Gly Tyr Ile Val Arg Asn Phe His Tyr
          35           40           45
Leu Leu Ile Lys Asp Gly Cys Arg Val Pro Ala Val Val Phe Thr Thr
          50           55           60
Leu Arg Gly Arg Gln Leu Cys Ala Pro Pro Asp Gln Pro Trp Val Glu
          65           70           75           80
Arg Ile Ile Gln Arg Leu Gln Arg Thr Ser Ala Lys Met Lys Arg Arg
          85           90           95
Ser Ser

```

&lt;210&gt; 274

&lt;211&gt; 684

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 274

```

cattcccagc ctcacatcac tcacaccttg catttcaccc ctgcatccca gtgcgcctgc 60
agcctcacac agatcctgca cacaccaga cagctggcgc tcacacattc accgttggcc 120
tgctctgtgt caccctccat ggccctgcta ctggccctca gcctgctggt tctctggact 180
tccccagccc caactctgag tggcaccaat gatgctgaag actgctgcct gtctgtgacc 240
cagaaaccca tccctgggta catcgtgagg aacttccact accttctcat caaggatggc 300
tgagggtgc ctgctgtagt gttcaccaca ctgaggggcc gccagctctg tgcaccccca 360
gaccagccct gggtagaacg catcatccag agactgcaga ggacctcagc caagatgaag 420
cgccgcagca gttaacctat gaccgtgcag agggagcccg gagtccgagt caagcattgt 480
gaattattac ctaacctggg gaaccgagga ccagaaggaa ggaccaggct tccagctcct 540
ctgcaccaga cctgaccagc caggacaggg cctggggtgt gtgtgagtgt gagtgtgagc 600
gagagggtga gtgtggtcag agtaaagctg ctccaccccc agattgcaat gctaccaata 660
aagccgcctg gtgtttacaa ct'aa 684

```

&lt;210&gt; 275

&lt;211&gt; 125

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 275

```

Met Lys Lys Ser Gly Val Leu Phe Leu Leu Gly Ile Ile Leu Leu Val
 1           5           10          15
Leu Ile Gly Val Gln Gly Thr Pro Val Val Arg Lys Gly Arg Cys Ser
          20           25           30
Cys Ile Ser Thr Asn Gln Gly Thr Ile His Leu Gln Ser Leu Lys Asp
          35           40           45
Leu Lys Gln Phe Ala Pro Ser Pro Ser Cys Glu Lys Ile Glu Ile Ile

```

50  
 Ala Thr Leu Lys Asn Gly Val Gln Thr Cys Leu Asn Pro Asp Ser Ala  
 65 70 75 80  
 Asp Val Lys Glu Leu Ile Lys Lys Trp Glu Lys Gln Val Ser Gln Lys  
 85 90 95  
 Lys Lys Gln Lys Asn Gly Lys Lys His Gln Lys Lys Lys Val Leu Lys  
 100 105 110  
 Val Arg Lys Ser Gln Arg Ser Arg Gln Lys Lys Thr Thr  
 115 120 125

&lt;210&gt; 276

&lt;211&gt; 2545

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 276

atccaataca ggagtgactt ggaactccat tctatcacta tgaagaaaag tgggtgttctt 60  
 ttccctcttgg gcatcatctt gctggttctg attggagtg c aaggaacccc agtagtgaga 120  
 aagggtcgtt gttcctgcat cagcaccaac caagggacta tccacctaca atccttgaaa 180  
 gaccttaaac aatttgcccc aagcccttcc tgcgagaaaa ttgaaatcat tgctacactg 240  
 aagaatggag ttcaaacatg tctaaaccca gattcagcag atgtgaagga actgattaaa 300  
 aagtgggaga aacagggtcag ccaaaaagaaa aagcaaaaaga atgggaaaaa acatcaaaaa 360  
 aagaaagttc tgaaagtctg aaaatctcaa cgttctcgtc aaaagaagac tacataagag 420  
 accacttcac caataagtat tctgtgttaa aaatgttcta ttttaattat accgctatca 480  
 ttccaaagga ggatggcata taatacaaaag gcttattaat ttgactagaa aatttaaaac 540  
 attactctga aattgtaact aaagttagaa agttgatttt aagaatccaa acgttaagaa 600  
 ttgttaaagg ctatgattgt ctttgttctt ctaccacca ccagttgaat ttcacatgc 660  
 ttaaggccat gatttttagca ataccatgt ctacacagat gtaccacca ccacatcca 720  
 ctcaaacag ctgcctggaa gagcagccct aggcctccac gtactgcagc ctccagagag 780  
 tatctgagga acatgtcagc aagtcctaag cctgttagca tgctgggtgag ccaagcagtt 840  
 tgaaattgag ctggacctca ccaagctgct gtggccatca acctctgtat ttgaatcagc 900  
 ctacaggcct cacacacaat gtgtctgaga gattcatgct gattgttatt ggggtatcacc 960  
 actggagatc accagtgtgt ggctttcaga gcctccttcc tggctttgga agccatgtga 1020  
 ttccatcttg cccgctcagg ctgaccactt tatttctttt tggtccctt tgcttcattc 1080  
 aagtcagctc ttctccatcc taccacaatg cagtgccttt ctctctcca gtgcacctgt 1140  
 catatgctct gatttatctg agtcaactcc tttctcatct tgtccccaac accccacaga 1200  
 agtgctttct tctcccaatt catcctcact cagtccagct tagttcaagt cctgcctctt 1260  
 aaataaacct ttttgacac acaaatatc ttaaaactcc tgtttcactt gggtcagtag 1320  
 cacatgggtg aacactcaat ggtaactaa ttcttggtg tttatcctat ctctccaacc 1380  
 agattgtcag ctccctgagg gcaagagcca cagtatatct ccctgtttct tccacagtgc 1440  
 ctaataatac tgtggaacta ggttttaata atttttaaat tgatgttgtt atgggcagga 1500  
 tggcaaccag accattgtct cagagcaggt gctggctctt tctggctac tccatgtttg 1560  
 ctagcctctg gtaacctctt acttattatc ttcaggacac tcaactacagg gaccaggat 1620  
 gatgcaacat ccttgtcttt ttatgacagg atgtttgctc agcttctcca acaataagaa 1680  
 gcacgtggta aaacacttgc ggatattctg gactgttttt aaaaaatata cagtttaccg 1740  
 aaaatcatat aatctttaca tgaaaaggac tttatagatc agccagtgc caaccttttc 1800  
 ccaaccatac aaaaattcct tttcccgaag gaaaagggtt ttctcaataa gcctcagctt 1860  
 tctaagatct aacaagatag ccaccgagat ccttatcgaa actcatttta ggcaaataatg 1920  
 agttttattg tccgtttact tgtttcagag tttgtattgt gattatcaat taccacacca 1980  
 tctcccatga agaaagggaa cggatgaagta ctaagcgcta gaggaagcag ccaagtcggt 2040  
 tagtgaagc atgattgggtg cccagttagc ctctgcagga tgtggaaacc tccttccagg 2100  
 ggaggttcag tgaattgtgt aggagaggtt gtctgtggcc agaatttaaa cctatactca 2160  
 ctttcccaaa ttgaatcact gctcacactg ctgatgattt agagtgtctg ccggtggaga 2220  
 tcccaccga acgtcttctc taatcatgaa actccctagt tccttcatgt aacttccctg 2280  
 aaaaatctaa gtgtttcata aatttgagag tctgtgacct acttaccttg catctcacag 2340  
 gtagacagta tataactaac aaccaagac tacatattgt cactgacaca caggtataa 2400  
 tcatttatca tatatataca tacatgcata cactctcaaa gcaaataatt tttcacttca 2460  
 aaacagtatt gacttgata ccttgtaatt tgaaatatct tctttgttaa aatagaatgg 2520  
 tatcaataaa tagaccatta atcag 2545



<210> 277  
 <211> 98  
 <212> PRT  
 <213> Homo sapiens

<400> 277  
 Met Asn Gln Thr Ala Ile Leu Ile Cys Cys Leu Ile Phe Leu Thr Leu  
 1 5 10 15  
 Ser Gly Ile Gln Gly Val Pro Leu Ser Arg Thr Val Arg Cys Thr Cys  
 20 25 30  
 Ile Ser Ile Ser Asn Gln Pro Val Asn Pro Arg Ser Leu Glu Lys Leu  
 35 40 45  
 Glu Ile Ile Pro Ala Ser Gln Phe Cys Pro Arg Val Glu Ile Ile Ala  
 50 55 60  
 Thr Met Lys Lys Lys Gly Glu Lys Arg Cys Leu Asn Pro Glu Ser Lys  
 65 70 75 80  
 Ala Ile Lys Asn Leu Leu Lys Ala Val Ser Lys Glu Met Ser Lys Arg  
 85 90 95  
 Ser Pro

<210> 278  
 <211> 1172  
 <212> DNA  
 <213> Homo sapiens

<400> 278  
 gagacattcc tcaattgctt agacatattc tgagcctaca gcagaggaac ctccagtctc 60  
 agcaccatga atcaaaactgc gattctgatt tgctgcctta tctttctgac tctaagtggc 120  
 attcaaggag tacctctctc tagaaccgta cgctgtacct gcatcagcat tagtaatcaa 180  
 cctgttaatc caaggtcttt agaaaaactt gaaattattc ctgcaagcca attttgcca 240  
 cgtgttgaga tcattgctac aatgaaaaag aagggtgaga agagatgtct gaatccagaa 300  
 tcgaaggcca tcaagaattt actgaaagca gttagcaagg aaatgtctaa aagatctcct 360  
 taaaaccaga ggggagcaaa atcgatgcag tgcttccaag gatggaccac acagaggctg 420  
 cctctcccat cacttcccta catggagtat atgtcaagcc ataattgttc ttagtttgca 480  
 gttacactaa aagggtgacca atgatggtca ccaaatcagc tgctactact cctgtaggaa 540  
 ggtaaatgtt catcatccta agctattcag taataactct accctggcac tataatgtaa 600  
 gctctactga ggtgctatgt tcttagtgga tgctctgacc ctgcttcaaa tatttccctc 660  
 acctttccca tcttccaagg gtactaagga atctttctgc tttgggggtt atcagaattc 720  
 tcagaatctc aaataactaa aaggatgca atcaaatctg ctttttaag aatgctcttt 780  
 actcatgga cttccactgc catcctccca aggggcccaa attccttcag tggctacct 840  
 catacaatc caaacacata caggaaggta gaaatatctg aaaatgtatg tgtaagtatt 900  
 cttatttgaat gaaagactgt acaaagtata agtcttagat gtatatattt cctatatattg 960  
 tttcagtgt catggaataa catgtaatta agtactatgt atcaatgagt aacaggaaaa 1020  
 ttttaaaaat acagatagat atatgctctg catgttacat aagataaatg tgctgaatgg 1080  
 ttttcaaaata aaaatgaggt actctcctgg aaatatgaag aaagactatc taaatgttga 1140  
 aagatcaaaa ggttaataaa gtaattataa ct 1172

<210> 279  
 <211> 166  
 <212> PRT  
 <213> Homo sapiens

<400> 279  
 Met Lys Tyr Thr Ser Tyr Ile Leu Ala Phe Gln Leu Cys Ile Val Leu  
 1 5 10 15  
 Gly Ser Leu Gly Cys Tyr Cys Gln Asp Pro Tyr Val Lys Glu Ala Glu  
 20 25 30  
 Asn Leu Lys Lys Tyr Phe Asn Ala Gly His Ser Asp Val Ala Asp Asn  
 35 40 45

Gly Thr Leu Phe Leu Gly Ile Leu Lys Asn Trp Lys Glu Glu Ser Asp  
 50 55 60  
 Arg Lys Ile Met Gln Ser Gln Ile Val Ser Phe Tyr Phe Lys Leu Phe  
 65 70 75 80  
 Lys Asn Phe Lys Asp Gln Ser Ile Gln Lys Ser Val Glu Thr Ile  
 85 90 95  
 Lys Glu Asp Met Asn Val Lys Phe Phe Asn Ser Asn Lys Lys Lys Arg  
 100 105 110  
 Asp Asp Phe Glu Lys Leu Thr Asn Tyr Ser Val Thr Asp Leu Asn Val  
 115 120 125  
 Gln Arg Lys Ala Ile His Glu Leu Ile Gln Val Met Ala Glu Leu Ser  
 130 135 140  
 Pro Ala Ala Lys Thr Gly Lys Arg Lys Arg Ser Gln Met Leu Phe Gln  
 145 150 155 160  
 Gly Arg Arg Ala Ser Gln  
 165

<210> 280  
 <211> 1193  
 <212> DNA  
 <213> Homo sapiens

<400> 280  
 tgaagatcag ctattagaag agaaagatca gttaagtcc tgggacctga tcagcttgat 60  
 acaagaacta ctgatttcaa cttctttggc ttaattctct cggaaacgat gaaatataca 120  
 agttatatct tggcttttca gctctgcac gttttgggtt ctcttggtg ttactgccag 180  
 gacccatatg taaaagaagc agaaaacctt aagaaatatt ttaatgcagg tcattcagat 240  
 gtagcggata atggaactct tttcttaggc attttgaaga attggaaaga ggagagtgc 300  
 agaaaaataa tgcagagcca aattgtctcc ttttacttca aactttttta aaacttttaa 360  
 gatgaccaga gcatccaaaa gagtgtggag accatcaagg aagacatgaa tgtcaagttt 420  
 ttcaatagca acaaaaagaa acgagatgac ttcgaaaagc tgactaatta ttcggtaact 480  
 gacttgaatg tccaacgcaa agcaatacat gaactcatcc aagtgatggc tgaactgtcg 540  
 ccagcagcta aaacagggaa gcgaaaaagg agtcagatgc tgtttcaagg tcgaagagca 600  
 tcccagtaat ggttgtcctg cctgcaatat ttgaatttta aatctaaatc tatttattaa 660  
 tatttaacat tatttatatg gggaatatat ttttagactc atcaatcaaa taagtattta 720  
 taatagcaac ttttgtgtaa tgaaaatgaa tatctattaa tatatgtatt atttataatt 780  
 cctatatcct gtgactgtct cacttaatcc tttgttttct gactaattag gcaaggctat 840  
 gtgattacaa ggctttatct caggggcca ctaggcagcc aacctaagca agatcccatg 900  
 ggttgtgtgt ttatttcact tgatgataca atgaacactt ataagtgaag tgatactatc 960  
 cagttactgc cggtttgaaa atatgcctgc aatctgagcc agtgctttta tggcatgtca: 1020  
 gacagaactt gaatgtgtca ggtgacctg atgaaaacat agcatctcag gagatttcat 1080  
 gcctggtgct tccaaatatt gttgacaact gtgactgtac ccaaatggaa agtaactcat 1140  
 ttgttaaat tatcaatatc taatatatat gaataaagtg taagttcaca act 1193

<210> 281  
 <211> 34  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer

<400> 281  
 gcggaatcat gggcaccaat gatgctgaag actg

34

<210> 282  
 <211> 34  
 <212> DNA  
 <213> Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 282

gcgggacccct taactgctgc ggcgcttcac ctg

34

&lt;210&gt; 283

&lt;211&gt; 35

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 283

gccgaattca cccagtagt gagaaagggt cgctg

35

&lt;210&gt; 284

&lt;211&gt; 39

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 284

cgcgatccct tatgtagtct tcttttgacg agaacgttg

39

&lt;210&gt; 285

&lt;211&gt; 36

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 285

gccgaattcg tacctctctc tagaaccgta cgctgt

36

&lt;210&gt; 286

&lt;211&gt; 40

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 286

gcgggacccct taaggagatc ttttagacat ttccttgcta

40

&lt;210&gt; 287

&lt;211&gt; 33

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 287

gcggaatcat gtgttactgc caggacccat atg

33

<210> 288  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer

<400> 288  
 gcgggatcct tactgggatg ctcttcgacc ttg

33

<210> 289  
 <211> 91  
 <212> PRT  
 <213> Homo sapiens

<400> 289  
 Met Lys Val Ser Ala Ala Ala Leu Ala Val Ile Leu Ile Ala Thr Ala  
 1 5 10 15  
 Leu Cys Ala Pro Ala Ser Ala Ser Pro Tyr Ser Ser Asp Thr Thr Pro  
 20 25 30  
 Cys Cys Phe Ala Tyr Ile Ala Arg Pro Leu Pro Arg Ala His Ile Lys  
 35 40 45  
 Glu Tyr Phe Tyr Thr Ser Gly Lys Cys Ser Asn Pro Ala Val Val Phe  
 50 55 60  
 Val Thr Arg Lys Asn Arg Gln Val Cys Ala Asn Pro Glu Lys Lys Trp  
 65 70 75 80  
 Val Arg Glu Tyr Ile Asn Ser Leu Glu Met Ser  
 85 90

<210> 290  
 <211> 1237  
 <212> DNA  
 <213> Homo sapiens

<400> 290  
 gctgcagagg attcctgcag aggatcaaga cagcacgtgg acctcgca ca gcctctccca 60  
 caggtagcat gaaggtctcc gcggcagccc tcgctgtcat cctcattgct actgcccctct 120  
 gcgctcctgc atctgcctcc ccatattcct cggacaccac accctgctgc tttgcctaca 180  
 ttgcccgcgc actgccccgt gccacatca aggagtattt ctacaccagt ggcaagtgtc 240  
 ccaaccagc agtcgtcttt gtcacccgaa agaaccgcca agtgtgtgcc aaccagaga 300  
 agaaatgggt tcgggagtac atcaactctt tggagatgag ctaggatgga gagtccctga 360  
 acctgaactt acacaaattt gcctgtttct gcttgccttt gtcctagctt gggaggcttc 420  
 cctcactat cctacccac ccgctccttg aagggccag attctaccac acagcagcag 480  
 ttacaaaaac cttcccagc ctggacgtgg tggctcacgc ctgtaatccc agcacttttg 540  
 gaggccagg tgggtggatc acttgagggtc aggagtgcga gaccagcctg gccaacatga 600  
 tgaaacccca tcttactaa aaatacaaaa aattagccgg gcgtggtagc gggcgccgtg 660  
 agtcccagct actcgggagg ctgaggcagg agaattggcgt gaaccgggga ggcgaggctt 720  
 gcagttagcc gagatcgcc cactgcactc cagcctgggc gacagagcga gactccgtct 780  
 caaaaaaaaa aaaaaaaaaa aaaatacaaa aattagccgg gcgtggtagc ccacgcctgt 840  
 aatcccagct actcgggagg ctaaggcagg aaaattgttt gaaccagga ggtggaggct 900  
 gcagttagct gagattgtgc cacttcactc cagcctgggt gacaaagtga gactccgtca 960  
 caacaacaac acaaaaaagc ttccccaaact aaagcctaga agagcttctg aggcgctgct 1020  
 ttgtcaaaag gaagtctcta ggttctgagc tctggctttg ccttggcttt gccagggtc 1080  
 tgtgaccagg aaggaaagtc gcattgcctct agaggcaagg aggggaggaa cactgcactc 1140  
 ttaagcttcc gccgtctcaa cccctcacag gagcttactg gcaaacatga aaaatcggct 1200  
 taccattaaa gttctcaatg caaccataaa aaaaaaa 1237

<210> 291  
 <211> 33

<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 291  
cgcgatccg tgcagtcctg ctccgcctac ggc

33

<210> 292  
<211> 39  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 292  
ccgaattcctt atgctggtag ttcaactatt tcaaagtag

39

<210> 293  
<211> 16  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> THAP antigenic peptide

<400> 293  
Ala Val Arg Arg Lys Asn Phe Lys Pro Thr Lys Tyr Ser Ser Ile Cys  
1 5 10 15

<210> 294  
<211> 16  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Control peptide

<400> 294  
Gln Val Glu Lys Leu Arg Lys Lys Leu Lys Thr Ala Gln Gln Arg Cys  
1 5 10 15

<210> 295  
<211> 33  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 295  
ccgaattcag gatggtgcag tcctgctccg cct

33

<210> 296  
<211> 41  
<212> DNA  
<213> Artificial Sequence

<220>  
 <223> Primer  
  
 <400> 296  
 gcgctctaga ttatgctggg acttcaacta tttcaaagta g 41  
  
 <210> 297  
 <211> 38  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Primer  
  
 <400> 297  
 gcgtctagaa tgagtgatgg aggggctcag gactgttg 38  
  
 <210> 298  
 <211> 38  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Primer  
  
 <400> 298  
 gggcgggcgc ctatggccct ttaggggtct gtgaccgc 38  
  
 <210> 299  
 <211> 35  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Primer  
  
 <400> 299  
 gcgctcgagc tgcacctggg ctttctctgc cctgg 35  
  
 <210> 300  
 <211> 35  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Primer  
  
 <400> 300  
 cgaagcttac tgtgctcctt ttatctctgc ccaag 35  
  
 <210> 301  
 <211> 420  
 <212> DNA  
 <213> Homo sapiens  
  
 <400> 301  
 cttctctgcc ctgggccaag cctgcccag cctctctgtc ctctgcctgc ccagctggac 60  
 atctctgggc ctctctggag accagtgggg tgggctgtgg gggcgtcata ttgccctggc 120  
 ttggcatccc tcttgtggct gtaccctcc cagcagcccc aggactagca agtccccgag 180  
 atgggggtgg ggacagtggg tgatgccaaa gggtgtgggg gcaggggcgg ggcaggagca 240

ggaagggtccc ctgagttccc tcaccttggg cagagataaa aggagcacag ttccaggcgg 300  
 ggctgagcta gggcgtagct gtgatttcag gggcacctct ggcggctgcc gtgatttgag 360  
 aatctcgggt ctcttggtg actgatcctg ggagactgtg gatgaataat gctgggtgagt 420

<210> 302

<211> 378

<212> PRT

<213> Homo sapiens

<400> 302

Met	Asp	Leu	Gly	Lys	Pro	Met	Lys	Ser	Val	Leu	Val	Val	Ala	Leu	Leu
1				5					10					15	
Val	Ile	Phe	Gln	Val	Cys	Leu	Cys	Gln	Asp	Glu	Val	Thr	Asp	Asp	Tyr
			20					25					30		
Ile	Gly	Asp	Asn	Thr	Thr	Val	Asp	Tyr	Thr	Leu	Phe	Glu	Ser	Leu	Cys
		35					40					45			
Ser	Lys	Lys	Asp	Val	Arg	Asn	Phe	Lys	Ala	Trp	Phe	Leu	Pro	Ile	Met
	50					55					60				
Tyr	Ser	Ile	Ile	Cys	Phe	Val	Gly	Leu	Leu	Gly	Asn	Gly	Leu	Val	Val
65					70				75					80	
Leu	Thr	Tyr	Ile	Tyr	Phe	Lys	Arg	Leu	Lys	Thr	Met	Thr	Asp	Thr	Tyr
			85					90						95	
Leu	Leu	Asn	Leu	Ala	Val	Ala	Asp	Ile	Leu	Phe	Leu	Leu	Thr	Leu	Pro
		100						105					110		
Phe	Trp	Ala	Tyr	Ser	Ala	Ala	Lys	Ser	Trp	Val	Phe	Gly	Val	His	Phe
		115					120					125			
Cys	Lys	Leu	Ile	Phe	Ala	Ile	Tyr	Lys	Met	Ser	Phe	Phe	Ser	Gly	Met
	130					135					140				
Leu	Leu	Leu	Leu	Cys	Ile	Ser	Ile	Asp	Arg	Tyr	Val	Ala	Ile	Val	Gln
145					150				155					160	
Ala	Val	Ser	Ala	His	Arg	His	Arg	Ala	Arg	Val	Leu	Leu	Ile	Ser	Lys
			165					170						175	
Leu	Ser	Cys	Val	Gly	Ile	Trp	Ile	Leu	Ala	Thr	Val	Leu	Ser	Ile	Pro
	180							185					190		
Glu	Leu	Leu	Tyr	Ser	Asp	Leu	Gln	Arg	Ser	Ser	Ser	Glu	Gln	Ala	Met
	195						200					205			
Arg	Cys	Ser	Leu	Ile	Thr	Glu	His	Val	Glu	Ala	Phe	Ile	Thr	Ile	Gln
	210				215						220				
Val	Ala	Gln	Met	Val	Ile	Gly	Phe	Leu	Val	Pro	Leu	Leu	Ala	Met	Ser
225					230				235					240	
Phe	Cys	Tyr	Leu	Val	Ile	Ile	Arg	Thr	Leu	Leu	Gln	Ala	Arg	Asn	Phe
			245						250					255	
Glu	Arg	Asn	Lys	Ala	Ile	Lys	Val	Ile	Ile	Ala	Val	Val	Val	Val	Phe
		260						265					270		
Ile	Val	Phe	Gln	Leu	Pro	Tyr	Asn	Gly	Val	Val	Leu	Ala	Gln	Thr	Val
	275						280					285			
Ala	Asn	Phe	Asn	Ile	Thr	Ser	Ser	Thr	Cys	Glu	Leu	Ser	Lys	Gln	Leu
	290					295					300				
Asn	Ile	Ala	Tyr	Asp	Val	Thr	Tyr	Ser	Leu	Ala	Cys	Val	Arg	Cys	Cys
305					310				315					320	
Val	Asn	Pro	Phe	Leu	Tyr	Ala	Phe	Ile	Gly	Val	Lys	Phe	Arg	Asn	Asp
			325						330					335	
Leu	Phe	Lys	Leu	Phe	Lys	Asp	Leu	Gly	Cys	Leu	Ser	Gln	Glu	Gln	Leu
		340						345					350		
Arg	Gln	Trp	Ser	Ser	Cys	Arg	His	Ile	Arg	Arg	Ser	Ser	Met	Ser	Val
	355						360						365		
Glu	Ala	Glu	Thr	Thr	Thr	Thr	Phe	Ser	Pro						
	370					375									

<210> 303  
 <211> 1136  
 <212> DNA  
 <213> Homo sapiens

<400> 303  
 atggacctgg ggaaccaat gaaaagcgtg ctgggtggtg ctctccttgt cattttccag 60  
 gtatgcctgt gtcaagatga ggtcacggac gattacatcg gagacaacac cacagtggat 120  
 acactttgtt cgagtccttg tgctccaaga aggacgtgcg gaactttaaa gcctgggttcc 180  
 tccctatcat gtactccatc atttgtttcg tgggcctact gggcaatggg ctggtcgtgt 240  
 tgacctatat ctatttcaag aggtcaaga ccatgaccga tacctacctg ctcaacctgg 300  
 cgggtggcaga catcctcttc ctccctgacc ttcccttctg ggcctacagc gcggccaagt 360  
 cctgggtctt cgggtgccac ttttgcaagc tcatctttgc catctacaag atgagcttct 420  
 tcagtggcat gctcctactt ctttgcatca gcattgaccg ctacgtggcc atcgccagg 480  
 ctgtctcagc tcaccgccac cgtgcccgcg tcttctcat cagcaagctg tctgtgtgg 540  
 gcatctggat actagccaca gtgctctcca tcccagagct cctgtacagt gacctccaga 600  
 ggagcagcag tgagcaagcg atgcgatgct ctctcatcac agagcatgtg gaggccttta 660  
 tcaccatcca ggtggcccag atggtgatcg gctttctggt cccctgctg gccatgagct 720  
 tctgttacct tgtcatcatc cgcaccctgc tccaggcagc caactttgag cgcaacaagg 780  
 ccatcaaggt gatcatcgct gtggtcgtgg tcttcatagt cttccagctg ccctacaatg 840  
 ggggtgtcct ggcccagacg gtggccaact tcaacatcac cagtagcacc tgtgagctca 900  
 gtaagcaact caacatcgcc tacgacgtca cctacagcct ggctgctgc cgctgctgcg 960  
 tcaacccttt cttgtacgcc ttcacggcg tcaagtccg caacgatctc ttcaagctct 1020  
 tcaaggacct gggctgcctc agccaggagc agctccggca gtggtcttcc tgcggcaca 1080  
 tccggcgctc ctccatgagt gtggaggccg agaccaccac caccttctcc ccatag 1136

<210> 304  
 <211> 368  
 <212> PRT  
 <213> Homo sapiens

<400> 304  
 Met Val Leu Glu Val Ser Asp His Gln Val Leu Asn Asp Ala Glu Val  
 1 5 10 15  
 Ala Ala Leu Leu Glu Asn Phe Ser Ser Ser Tyr Asp Tyr Gly Glu Asn  
 20 25 30  
 Glu Ser Asp Ser Cys Cys Thr Ser Pro Pro Cys Pro Gln Asp Phe Ser  
 35 40 45  
 Leu Asn Phe Asp Arg Ala Phe Leu Pro Ala Leu Tyr Ser Leu Leu Phe  
 50 55 60  
 Leu Leu Gly Leu Leu Gly Asn Gly Ala Val Ala Ala Val Leu Leu Ser  
 65 70 75 80  
 Arg Arg Thr Ala Leu Ser Ser Thr Asp Thr Phe Leu Leu His Leu Ala  
 85 90 95  
 Val Ala Asp Thr Leu Leu Val Leu Thr Leu Pro Leu Trp Ala Val Asp  
 100 105 110  
 Ala Ala Val Gln Trp Val Phe Gly Ser Gly Leu Cys Lys Val Ala Gly  
 115 120 125  
 Ala Leu Phe Asn Ile Asn Phe Tyr Ala Gly Ala Leu Leu Leu Ala Cys  
 130 135 140  
 Ile Ser Phe Asp Arg Tyr Leu Asn Ile Val His Ala Thr Gln Leu Tyr  
 145 150 155 160  
 Arg Arg Gly Pro Pro Ala Arg Val Thr Leu Thr Cys Leu Ala Val Trp  
 165 170 175  
 Gly Leu Cys Leu Leu Phe Ala Leu Pro Asp Phe Ile Phe Leu Ser Ala  
 180 185 190  
 His His Asp Glu Arg Leu Asn Ala Thr His Cys Gln Tyr Asn Phe Pro  
 195 200 205  
 Gln Val Gly Arg Thr Ala Leu Arg Val Leu Gln Leu Val Ala Gly Phe  
 210 215 220  
 Leu Leu Pro Leu Leu Val Met Ala Tyr Cys Tyr Ala His Ile Leu Ala



225		230		235		240									
Val	Leu	Leu	Val	Ser	Arg	Gly	Gln	Arg	Arg	Leu	Arg	Ala	Met	Arg	Leu
				245				250						255	
Val	Val	Val	Val	Val	Val	Ala	Phe	Ala	Leu	Cys	Trp	Thr	Pro	Tyr	His
				260				265					270		
Leu	Val	Val	Leu	Val	Asp	Ile	Leu	Met	Asp	Leu	Gly	Ala	Leu	Ala	Arg
			275				280					285			
Asn	Cys	Gly	Arg	Glu	Ser	Arg	Val	Asp	Val	Ala	Lys	Ser	Val	Thr	Ser
	290					295					300				
Gly	Leu	Gly	Tyr	Met	His	Cys	Cys	Leu	Asn	Pro	Leu	Leu	Tyr	Ala	Phe
305					310					315					320
Val	Gly	Val	Lys	Phe	Arg	Glu	Arg	Met	Trp	Met	Leu	Leu	Leu	Arg	Leu
			325					330						335	
Gly	Cys	Pro	Asn	Gln	Arg	Gly	Leu	Gln	Arg	Gln	Pro	Ser	Ser	Ser	Arg
			340					345					350		
Arg	Asp	Ser	Ser	Trp	Ser	Glu	Thr	Ser	Glu	Ala	Ser	Tyr	Ser	Gly	Leu
		355					360						365		

&lt;210&gt; 305

&lt;211&gt; 1670

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 305

```

ccaaccacaa gcaccaaagc agaggggagc gcagcacacc acccagcagc cagagcacca 60
gccagccat ggtccttgag gtgagtgacc accaagtgct aaatgacgcc gaggttgccg 120
ccctcctgga gaacttcagc tcttcctatg actatggaga aaacgagagt gactcgtgct 180
gtacctcccc gccctgcccc caggacttca gcctgaactt cgaccgggcc ttcctgccag 240
ccctctacag cctcctcttt ctgctggggc tgctgggcaa cggcgcggtg gcagccgtgc 300
tgctgagccg gcggacagcc ctgagcagca ccgacacctt cctgctccac ctagtctgtag 360
cagacacgct gctgggtgctg aactgcccgc tctgggcagt ggacgctgcc gtccagtggg 420
tctttggctc tggcctctgc aaagtggcag gtgccctctt caacatcaac ttctacgcag 480
gagccctcct gctggcctgc atcagctttg accgctacct gaacatagtt catgccaccc 540
agctctaccg ccggggggccc ccggcccgcg tgacctcac ctgcctggct gtctgggggc 600
tctgcctgct tttcgccctc ccagacttca tcttcctgtc ggcccaccac gacgagcgcc 660
tcaacgccac cactgccaa tacaacttcc cacaggtggg ccgcacggct ctgcgggtgc 720
tgcaagtggg ggctggcttt ctgctgcccc tgctggtcac ggctactgct tatgccaca 780
tcctggccgt gctgctgggt tccagggggc agcggcgccct gcgggccatg cggctgggtg 840
tggtggctgt ggtggccttt gccctctgct ggacccccca tcacctgggt gtgctgggtg 900
acatcctcat ggacctgggc gctttggccc gcaactgtgg ccgagaaagc agggtagacg 960
tggccaagtc ggtcacctca ggcctgggct acatgcactg ctgcctcaac ccgctgctct 1020
atgcctttgt aggggtcaag ttccgggagc ggatgtggat gctgctcttg cgcctgggct 1080
gccccaaaca gagagggctc cagaggcagc catcgtcttc ccgcccggat tcatcctggg 1140
ctgagacctc agaggcctcc tactcgggct tgtgaggccg gaatccgggc tcccctttcg 1200
cccacagtct gacttccccg cattccaggc tctccctccc ctctgccggc tctggctctc 1260
cccaatatcc tcgctcccgg gactcactgg cagccccagc accaccaggt ctcccgggaa 1320
gccacctccc cagctctgag gactgcacca ttgctgctcc ttagctgcca agccccatcc 1380
tgccgcccga ggtggctgcc tggagcccca ctgcccttct catttggaac ctaaaacttc 1440
atcttcccca agtgccgggga gtacaaggca tggcgtagag ggtgctgccc catgaagcca 1500
cagcccaggc ctccagctca gcagtgactg tggccatggg cccaagacc tctatatattg 1560
ctcttttatt tttatgtcta aaatcctgct taaaactttt caataaacia gatcgtcagg 1620
acaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1670

```

&lt;210&gt; 306

&lt;211&gt; 11

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; THRE consensus sequence

<221> misc\_feature  
<222> (1)...(1)  
<223> N=A or T

<221> misc\_feature  
<222> (2)...(2)  
<223> N=G, C or A

<221> misc\_feature  
<222> (4)...(4)  
<223> N= A or G

<221> misc\_feature  
<222> (5)...(5)  
<223> N=T, C or A

<221> misc\_feature  
<222> (6)...(6)  
<223> N= G or T

<221> misc\_feature  
<222> (11)...(11)  
<223> N= A, T, G or C

<400> 306  
nntnnnggca n

<210> 307  
<211> 11  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> THRE variant one

<400> 307  
agtaaggga a

<210> 308  
<211> 11  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Mutant THRE

<400> 308  
agtaatttca a

<210> 309  
<211> 11  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Mutant THRE

<400> 309  
agtaagggtca a

11

11

11

11

<210> 310  
<211> 11  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Mutant THRE

<400> 310  
agtaagtgc a

11

<210> 311  
<211> 11  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Mutant THRE

<400> 311  
agtaagggcc a

11

<210> 312  
<211> 11  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Mutant THRE

<400> 312  
agtaagggaa a

11

<210> 313  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Sequence containing THRE variant one

<400> 313  
agcaagtaag ggcaaactac ttcac

25

<210> 314  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Sequence containing THRE mutant

<400> 314  
agcaagtaat ttcaaactac ttcac

25

<210> 315  
<211> 319  
<212> DNA  
<213> Homo sapiens

<400> 315

```

cgtcgctggg tgcaccgcga ccacgggcag agccacgcgg cgggaggact acaactcccg 60
gcacaccccg cgccgccccg cctctactcc cagaaggccg cggggggtgg accgcctaag 120
agggcggtgcg ctcccgacat gccccgcggc gcgccattaa ccgccagatt tgaatcgcg 180
gacccggttg cagaggtggc ggccggcgga tgggtgcccc gacgttgccc cctgcctggc 240
agccctttct caaggaccac cgcactctta cattcaagaa ctggcccttc ttggagggt 300
gcgcctgcac cccggagcg                                     319

```

&lt;210&gt; 316

&lt;211&gt; 17

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; DR-5-type element

&lt;221&gt; misc\_feature

&lt;222&gt; (7)...(7)

&lt;223&gt; N=A, T, G, or C

&lt;221&gt; misc\_feature

&lt;222&gt; (8)...(8)

&lt;223&gt; N=A, T, G, or C

&lt;221&gt; misc\_feature

&lt;222&gt; (9)...(9)

&lt;223&gt; N=A, T, G, or C

&lt;221&gt; misc\_feature

&lt;222&gt; (10)...(10)

&lt;223&gt; N=A, T, G, or C

&lt;400&gt; 316

gggcaannnn ngggcac

17

&lt;210&gt; 317

&lt;211&gt; 600

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 317

```

tgggtgccgtg cgccgtgtagt cccagctact cgggagggtg aggcagaaga atgcactcca 60
gcctgggcga cagagggata ttccgtctca aacaaacaaa aaatcactcg ctgcgttttt 120
tattctgaca tgggtgcagga aggtaaattc aagacaactt aggtactcag ttttagaagt 180
cgacaggaca gaattacgga aacaaattta agcgttcccc cttttagctc caaatataat 240
gtgttccaga aaggtaacca tctaggaaac tccaaggctc agaccaccac cggatgcccc 300
cacttcagga gcatttatat aacttcgtgg ttatgtcaga gacgagaaaa ccattgaca 360
accaaaccac taaaccgaa catccggcgc aagccgcacg caggcgaga tttactagcg 420
tcagagccga tgggtcccggt aggtgggggt ggggtggtgg tggcctagcc acttcccata 480
atgccgcgtt ccggaagtta ttgctttcca ggggtcactc tggcttcgac tccgtcgctc 540
tcaattcgtc accaggagga agacggagct ggctgccag cccaaaggcc catgagggga 600

```

&lt;210&gt; 318

&lt;211&gt; 11

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; THRE-type element

&lt;400&gt; 318

agtgtgggca t

11

&lt;210&gt; 319

&lt;211&gt; 25

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Sequence containing THRE mutant

&lt;400&gt; 319

agcaagtaag gtcaaactac ttcac

25

&lt;210&gt; 320

&lt;211&gt; 25

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Sequence containing THRE mutant

&lt;400&gt; 320

agcaagtaag tgcaaactac ttcac

25

&lt;210&gt; 321

&lt;211&gt; 25

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Sequence containing THRE mutant

&lt;400&gt; 321

agcaagtaag ggccaactac ttcac

25

&lt;210&gt; 322

&lt;211&gt; 25

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Sequence containing THRE mutant

&lt;400&gt; 322

agcaagtaag ggaaaactac ttcac

25

&lt;210&gt; 323

&lt;211&gt; 94

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 323

Met	Ser	Val	Lys	Gly	Met	Ala	Ile	Ala	Leu	Ala	Val	Ile	Leu	Cys	Ala
1				5					10					15	
Thr	Val	Val	Gln	Gly	Phe	Pro	Met	Phe	Lys	Arg	Gly	Arg	Cys	Leu	Cys
			20					25					30		
Ile	Gly	Pro	Gly	Val	Lys	Ala	Val	Lys	Val	Ala	Asp	Ile	Glu	Lys	Ala
			35				40					45			
Ser	Ile	Met	Tyr	Pro	Ser	Asn	Asn	Cys	Asp	Lys	Ile	Glu	Val	Ile	Ile
	50					55				60					
Thr	Leu	Lys	Glu	Asn	Lys	Gly	Gln	Arg	Cys	Leu	Asn	Pro	Lys	Ser	Lys

BNSDOCID: <WO\_\_\_\_\_2004055050A2 | >

<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 327  
ccgaattccc accatgaaga aaagtgggtgt tcttt 35

<210> 328  
<211> 35  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 328  
ccggatcctg tagtcttctt ttgacgagaa cgttg 35

<210> 329  
<211> 34  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 329  
cctctagacc accatgggcc ttgaggtgag tgac 34

<210> 330  
<211> 32  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 330  
cccgcggccg ctcacaagcc cgagtaggag gc 32

<210> 331  
<211> 34  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 331  
ggcattcaat tgctcgagtt taaacgcggc cgcg 34

<210> 332  
<211> 38  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 332  
aatccgcggc cgcgtttaa ctcgagcaat tgaatgcc 38

<210> 333  
<211> 26  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 333  
atgaagtgcc ttttgactt agcctt 26

<210> 334  
<211> 33  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 334  
tcataaaaat taaaaactca aatataattg agg 33

<210> 335  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 335  
atgggccaga ctgttaccac tc 22

<210> 336  
<211> 16  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 336  
ttagggggcc tcgcgg 16

<210> 337  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 337  
atggtgcagt cctgctccgc 20

<210> 338  
<211> 31  
<212> DNA



<213> Artificial Sequence

<220>

<223> Primer

<400> 338

gccaatgtt atgctggtac ttcaactatt t

31

<210> 339

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 339

atgagtgatg gaggggctca gg

22

<210> 340

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 340

ggaattccta tggcccttta ggg

23

<210> 341

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 341

atgaccccag tagtgagaaa gggtc

25

<210> 342

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 342

ggaattctta tgtagtcttc ttttgacgag a

31

<210> 343

<211> 142

<212> PRT

<213> Homo sapiens

<400> 343

Met Gly Ala Pro Thr Leu Pro Pro Ala Trp Gln Pro Phe Leu Lys Asp

1

5

10

15

His Arg Ile Ser Thr Phe Lys Asn Trp Pro Phe Leu Glu Gly Cys Ala

20 25 30  
 Cys Thr Pro Glu Arg Met Ala Glu Ala Gly Phe Ile His Cys Pro Thr  
 35 40 45  
 Glu Asn Glu Pro Asp Leu Ala Gln Cys Phe Phe Cys Phe Lys Glu Leu  
 50 55 60  
 Glu Gly Trp Glu Pro Asp Asp Asp Pro Ile Glu Glu His Lys Lys His  
 65 70 75 80  
 Ser Ser Gly Cys Ala Phe Leu Ser Val Lys Lys Gln Phe Glu Glu Leu  
 85 90 95  
 Thr Leu Gly Glu Phe Leu Lys Leu Asp Arg Glu Arg Ala Lys Asn Lys  
 100 105 110  
 Ile Ala Lys Glu Thr Asn Asn Lys Lys Lys Glu Phe Glu Glu Thr Ala  
 115 120 125  
 Lys Lys Val Arg Arg Ala Ile Glu Gln Leu Ala Ala Met Asp  
 130 135 140

&lt;210&gt; 344

&lt;211&gt; 1619

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 344

ccgccagatt tgaatcgcg gacccggttg cagaggtggc ggcggcgcca tgggtgcccc 60  
 gacgttgccc cctgcctggc agccctttct caaggaccac cgcattctcta cattcaagaa 120  
 ctggcccttc ttggagggct ggcctgcac cccggagcgg atggccgagg ctggcttcat 180  
 cactgcccc actgagaacg agccagactt ggcccagtgt ttcttctgct tcaaggagct 240  
 ggaaggctgg gagccagatg acgaccccat agaggaacat aaaaagcatt cgtccggttg 300  
 cgctttcctt tctgtcaaga agcagtttga agaattaacc cttggtgaat ttttgaaact 360  
 ggacagagaa agagccaaga acaaaattgc aaaggaaacc aacaataaga agaaagaatt 420  
 tgaggaaact gcgaagaaag tgcgccgtgc catcgagcag ctggctgcca tggattgagg 480  
 cctctggccg gagctgcctg gtcccagagt ggctgcacca cttccagggt ttattccctg 540  
 gtgccaccag ccttcctgtg ggccccttag caatgtctta ggaaaggaga tcaacatttt 600  
 caaattagat gtttcaactg tgctcctgtt ttgtcttgaa agtggcacca gaggtgcttc 660  
 tgcctgtgca gcgggtgctg ctggtaacag tggtctcttc tctctctctc tctctttttt 720  
 gggggctcat ttttgctgtt ttgattcccg ggcttaccag gtgagaagtg agggaggaag 780  
 aaggcagtgt cctttttgct agagctgaca gctttgttcg cgtgggcaga gccttcaca 840  
 gtgaatgtgt ctggacctca tgttgttgag gctgtcacag tcctgagtgt ggacttgga 900  
 ggtgcctgtt gaatctgagc tgcaggttcc ttatctgtca cacctgtgcc tcctcagagg 960  
 acagtttttt tgttgttgtg tttttttgtt tttttttttt ggtagatgca tgacttgtgt 1020  
 gtgatgagag aatggagaca gagtccctgg ctccctact gtttaacaac atggctttct 1080  
 tattttgttt gaattgttaa ttcacagaat agcacaact acaattaaaa ctaagcaca 1140  
 agccattcta agtcattggg gaaacggggg gaacttcagg tggatgagga gacagaatag 1200  
 agtgatagga agcgtctggc agatactcct tttgccactg ctgtgtgatt agacaggccc 1260  
 agtgagccgc ggggcacatg ctggccgctc ctccctcaga aaaaggcagt ggcctaaatc 1320  
 ctttttaaat gacttggtc gatgctgtgg gggactggct gggctgctgc aggcctgtgt 1380  
 tctgtcagcc caaccttcac atctgtcacg ttctccacac gggggagaga cgcagtccgc 1440  
 ccaggtcccc gctttctttg gaggcagcag ctcccgagc gctgaagtct ggcgtaagat 1500  
 gatggatttg attcgccctc ctccctgtca tagagctgca ggggtgattt ttacagcttc 1560  
 gctggaaacc tctggaggtc atctcggtg ttccctgagaa ataaaaagcc tgtcatttc 1619

&lt;210&gt; 345

&lt;211&gt; 202

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 345

Met Ala Thr Leu Ile Tyr Val Asp Lys Glu Asn Gly Glu Pro Gly Thr  
 1 5 10 15  
 Arg Val Val Ala Lys Asp Gly Leu Lys Leu Gly Ser Gly Pro Ser Ile  
 20 25 30

Lys Ala Leu Asp Gly Arg Ser Gln Val Ser Thr Pro Arg Phe Gly Lys  
 35 40 45  
 Thr Phe Asp Ala Pro Pro Ala Leu Pro Lys Ala Thr Arg Lys Ala Leu  
 50 55 60  
 Gly Thr Val Asn Arg Ala Thr Glu Lys Ser Val Lys Thr Lys Gly Pro  
 65 70 75 80  
 Leu Lys Gln Lys Gln Pro Ser Phe Ser Ala Lys Lys Met Thr Glu Lys  
 85 90 95  
 Thr Val Lys Ala Lys Ser Ser Val Pro Ala Ser Asp Asp Ala Tyr Pro  
 100 105 110  
 Glu Ile Glu Lys Phe Phe Pro Phe Asn Pro Leu Asp Phe Glu Ser Phe  
 115 120 125  
 Asp Leu Pro Glu Glu His Gln Ile Ala His Leu Pro Leu Ser Gly Val  
 130 135 140  
 Pro Leu Met Ile Leu Asp Glu Glu Arg Glu Leu Glu Lys Leu Phe Gln  
 145 150 155 160  
 Leu Gly Pro Pro Ser Pro Val Lys Met Pro Ser Pro Pro Trp Glu Ser  
 165 170 175  
 Asn Leu Leu Gln Ser Pro Ser Ser Ile Leu Ser Thr Leu Asp Val Glu  
 180 185 190  
 Leu Pro Pro Val Cys Cys Asp Ile Asp Ile  
 195 200

<210> 346  
 <211> 727  
 <212> DNA  
 <213> Homo sapiens

<400> 346  
 cggcctcaga tgaatgcggc tgtaagacc tgcaataatc cagaatggct actctgatct 60  
 atgttgataa ggaaaatgga gaaccaggca cccgtgtggt tgctaaggat gggctgaagc 120  
 tggggtctgg accttcaatc aaagccttag atgggagatc tcaagtttca acaccacgtt 180  
 ttggcaaaac gttagatgcc ccaccagcct tacctaaagc tactagaaag gctttgggaa 240  
 ctgtcaacag agctacagaa aagtctgtaa agaccaaggg acccctcaaa caaaaacagc 300  
 caagcttttc tgccaaaaag atgactgaga agactgttaa agcaaaaagc tctgttcctg 360  
 cctcagatga tgcctatcca gaaatagaaa aattctttcc cttcaatcct ctgactttg 420  
 agagttttga cctgcctgaa gagcaccaga ttgcgcacct ccccttgagt ggagtgcctc 480  
 tcatgatcct tgacgaggag agagagcttg aaaagctgtt tcagctgggc ccccttcac 540  
 ctgtgaagat gccctctcca ccatgggaat ccaatctgtt gcagtctcct tcaagcattc 600  
 tgtcgaccct ggatgttgaa ttgccacctg tttgctgtga catagatatt taaatttctt 660  
 agtgcttcag agtttgtgtg tatttgtatt aataaagcat tcttcaacag aaaaaaaaaa 720  
 aaaaaaa 727

<210> 347  
 <211> 79  
 <212> PRT  
 <213> Homo sapiens

<400> 347  
 Met Ser His Lys Gln Ile Tyr Tyr Ser Asp Lys Tyr Asp Asp Glu Glu  
 1 5 10 15  
 Phe Glu Tyr Arg His Val Met Leu Pro Lys Asp Ile Ala Lys Leu Val  
 20 25 30  
 Pro Lys Thr His Leu Met Ser Glu Trp Arg Asn Leu Gly Val  
 35 40 45  
 Gln Gln Ser Gln Gly Trp Val His Tyr Met Ile His Glu Pro Glu Pro  
 50 55 60  
 His Ile Leu Leu Phe Arg Arg Pro Leu Pro Lys Lys Pro Lys Lys  
 65 70 75

<210> 348  
 <211> 717  
 <212> DNA  
 <213> Homo sapiens

<400> 348

```

agagcgatca tgtcgcacaa acaaattttac tattcggaca aatacgacga cgaggagttt 60
gagtatcgac atgtcatgct gcccaaggac atagccaagc tggtccttaa aaccatctg 120
atgtctgaat ctgaatggag gaatcttggc gttcagcaga gtcagggatg ggtccattat 180
atgatccatg aaccagaacc tcacatcttg ctgttccggc gccactacc caagaaacca 240
aagaaatgaa gctggcaagc tacttttcag cctcaagctt tacacagctg tccttacttc 300
ctaacatctt tctgataaca ttattatggt gccttcttgt ttctcacttt gatattttaaa 360
agatgttcaa tacactgttt gaatgtgctg gtaactgctt tgcttcttga gtagagccac 420
caccaccata gccagccag atgagtgtct tgtggacca cagcctaagc tgagtgtgac 480
cccagaagcc acgatgtgct ctgtatccag aacacacttg gcagatggag gaagcatctg 540
agtttgagac catggctgtt acagggatca tgtaaacttg ctgtttttgt tttttctgcc 600
gggtgttgta tgtgtggtga cttgcggatt tatgtttcag tgtactggaa actttccatt 660
ttattcaaga aatctgttca tgttaaaagc cttgattaaa gaggaagttt ttataat 717

```

<210> 349  
 <211> 205  
 <212> PRT  
 <213> Homo sapiens

<400> 349

```

Met Ala Leu Gln Leu Ser Arg Glu Gln Gly Ile Thr Leu Arg Gly Ser
  1           5           10           15
Ala Glu Ile Val Ala Glu Phe Phe Ser Phe Gly Ile Asn Ser Ile Leu
      20           25           30
Tyr Gln Arg Gly Ile Tyr Pro Ser Glu Thr Phe Thr Arg Val Gln Lys
      35           40           45
Tyr Gly Leu Thr Leu Leu Val Thr Thr Asp Leu Glu Leu Ile Lys Tyr
      50           55           60
Leu Asn Asn Val Val Glu Gln Leu Lys Asp Trp Leu Tyr Lys Cys Ser
      65           70           75           80
Val Gln Lys Leu Val Val Val Ile Ser Asn Ile Glu Ser Gly Glu Val
      85           90           95
Leu Glu Arg Trp Gln Phe Asp Ile Glu Cys Asp Lys Thr Ala Lys Asp
      100          105          110
Asp Ser Ala Pro Arg Glu Lys Ser Gln Lys Ala Ile Gln Asp Glu Ile
      115          120          125
Arg Ser Val Ile Arg Gln Ile Thr Ala Thr Val Thr Phe Leu Pro Leu
      130          135          140
Leu Glu Val Ser Cys Ser Phe Asp Leu Leu Ile Tyr Thr Asp Lys Asp
      145          150          155          160
Leu Val Val Pro Glu Lys Trp Glu Glu Ser Gly Pro Gln Phe Ile Thr
      165          170          175
Asn Ser Glu Glu Val Arg Leu Arg Ser Phe Thr Thr Thr Ile His Lys
      180          185          190
Val Asn Ser Met Val Ala Tyr Lys Ile Pro Val Asn Asp
      195          200          205

```

<210> 350  
 <211> 1390  
 <212> DNA  
 <213> Homo sapiens

<400> 350

```

gggaagtgtc gttggagccg ctgtggttgc tgtccgcgga gtggaagcgc gtgcttttgt 60

```

```

ttgtgtccct ggccatggcg ctgcagctct cccgggagca gggaatcacc ctgcgcggga 120
gcgccgaaat cgtggccgag ttctttctcat tcggcatcaa cagcatttta tatcagcgtg 180
gcataatatcc atctgaaacc ttactcgcag tgcagaaata cggactcacc ttgcttgtaa 240
ctactgatct tgagctcata aaatacctaa ataattgtgt ggacaactg aaagattggt 300
tatacaagtg ttcagttcag aaactggttg tagttatctc aaatattgaa agtggtgagg 360
tcctggaaag atggcagttt gatattgagt gtgacaagac tgcaaaagat gacagtgcac 420
ccagagaaaa gtctcagaaa gctatccagg atgaaatccg ttcagtgatc agacagatca 480
cagctacggt gacattttctg ccactgttgg aagtttcttg ttcatttgat ctgctgattt 540
atacagacaa agatttgggt gtacctgaaa aatgggaaga gtcgggacca cagtttatta 600
ccaattctga ggaagtccgc cttcgttcac ttactactac aatccacaaa gtaaatagca 660
tggtggccta caaaattcct gtcaatgact gaggatgaca tgaggaaaat aatgtaattg 720
taattttgaa atgtggtttt cctgaaatca ggtcatctat agttgatatg ttttatttca 780
ttggttaatt ttacatgga gaaaaccaa atgatactta ctgaactgtg tgtaattggt 840
cctttatttt tttggtacct atttgactta ccatggagtt aacatcatga atttattgca 900
cattgttcaa aaggaaccag gaggtttttt tgtcaacatt gtgatgtata ttcctttgaa 960
gatagtaact gtagatggaa aaacttggtc tataaagcta gatgctttcc taaatcagat 1020
gttttgggtc agtagtttga ctcagtatag gtaggagat atttaagat aaaatacaac 1080
aaaggaagtc taaatattca gaatctttgt taagggtcctg aaagtaactc ataactata 1140
aacaatgaaa tattgctgta tagctccttt tgaccttcac ttcatgtata gttttcccta 1200
ttgaatcagt ttccaattat ttgactttaa tttatgtaac ttgaacctat gaagcaatgg 1260
atatttgtag tgtttaatgt tctgtgatac agaactctta aaaatgtttt ttcattgtgt 1320
ttataaaatc aagttttaag tgaaagttag gaaataaagt taagtttggt ttaaaaaaaaa 1380
aaaaaaaaa 1390

```

<210> 351  
 <211> 823  
 <212> PRT  
 <213> Homo sapiens

<400> 351

Met	Gly	Lys	Lys	Arg	Thr	Lys	Gly	Lys	Thr	Val	Pro	Ile	Asp	Asp	Ser
1				5					10					15	
Ser	Glu	Thr	Leu	Glu	Pro	Val	Cys	Arg	His	Ile	Arg	Lys	Gly	Leu	Glu
			20					25					30		
Gln	Gly	Asn	Leu	Lys	Lys	Ala	Leu	Val	Asn	Val	Glu	Trp	Asn	Ile	Cys
		35					40					45			
Gln	Asp	Cys	Lys	Thr	Asp	Asn	Lys	Val	Lys	Asp	Lys	Ala	Glu	Glu	Glu
	50				55					60					
Thr	Glu	Glu	Lys	Pro	Ser	Val	Trp	Leu	Cys	Leu	Lys	Cys	Gly	His	Gln
65					70				75					80	
Gly	Cys	Gly	Arg	Asn	Ser	Gln	Glu	Gln	His	Ala	Leu	Lys	His	Tyr	Leu
			85					90						95	
Thr	Pro	Arg	Ser	Glu	Pro	His	Cys	Leu	Val	Leu	Ser	Leu	Asp	Asn	Trp
			100					105					110		
Ser	Val	Trp	Cys	Tyr	Val	Cys	Asp	Asn	Glu	Val	Gln	Tyr	Cys	Ser	Ser
		115				120					125				
Asn	Gln	Leu	Gly	Gln	Val	Val	Asp	Tyr	Val	Arg	Lys	Gln	Ala	Ser	Ile
	130				135						140				
Thr	Thr	Pro	Lys	Pro	Ala	Glu	Lys	Asp	Asn	Gly	Asn	Ile	Glu	Leu	Glu
145					150				155					160	
Asn	Lys	Lys	Leu	Glu	Lys	Glu	Ser	Lys	Asn	Glu	Gln	Glu	Arg	Glu	Lys
			165					170						175	
Lys	Glu	Asn	Met	Ala	Lys	Glu	Asn	Pro	Pro	Met	Asn	Ser	Pro	Cys	Gln
		180						185					190		
Ile	Thr	Val	Lys	Gly	Leu	Ser	Asn	Leu	Gly	Asn	Thr	Cys	Phe	Phe	Asn
	195					200						205			
Ala	Val	Met	Gln	Asn	Leu	Ser	Gln	Thr	Pro	Val	Leu	Arg	Glu	Leu	Leu
	210				215						220				
Lys	Glu	Val	Lys	Met	Ser	Gly	Thr	Ile	Val	Lys	Ile	Glu	Pro	Pro	Asp
225					230					235				240	
Leu	Ala	Leu	Thr	Glu	Pro	Leu	Glu	Ile	Asn	Leu	Glu	Pro	Pro	Gly	Pro

				245					250				255				
Leu	Thr	Leu	Ala	Met	Ser	Gln	Phe	Leu	Asn	Glu	Met	Gln	Glu	Thr	Lys		
			260					265					270				
Lys	Gly	Val	Val	Thr	Pro	Lys	Glu	Leu	Phe	Ser	Gln	Val	Cys	Lys	Lys		
		275					280					285					
Ala	Val	Arg	Phe	Lys	Gly	Tyr	Gln	Gln	Gln	Asp	Ser	Gln	Glu	Leu	Leu		
	290					295				300							
Arg	Tyr	Leu	Leu	Asp	Gly	Met	Arg	Ala	Glu	Glu	His	Gln	Arg	Val	Ser		
305					310					315					320		
Lys	Gly	Ile	Leu	Lys	Ala	Phe	Gly	Asn	Ser	Thr	Glu	Lys	Leu	Asp	Glu		
				325					330					335			
Glu	Leu	Lys	Asn	Lys	Val	Lys	Asp	Tyr	Glu	Lys	Lys	Lys	Ser	Met	Pro		
			340					345					350				
Ser	Phe	Val	Asp	Arg	Ile	Phe	Gly	Gly	Glu	Leu	Thr	Ser	Met	Ile	Met		
		355					360					365					
Cys	Asp	Gln	Cys	Arg	Thr	Val	Ser	Leu	Val	His	Glu	Ser	Phe	Leu	Asp		
	370					375				380							
Leu	Ser	Leu	Pro	Val	Leu	Asp	Asp	Gln	Ser	Gly	Lys	Lys	Ser	Val	Asn		
385					390					395					400		
Asp	Lys	Asn	Leu	Lys	Lys	Thr	Val	Glu	Asp	Glu	Asp	Gln	Asp	Ser	Glu		
			405					410						415			
Glu	Glu	Lys	Asp	Asn	Asp	Ser	Tyr	Ile	Lys	Glu	Arg	Ser	Asp	Ile	Pro		
			420					425					430				
Ser	Gly	Thr	Ser	Lys	His	Leu	Gln	Lys	Lys	Ala	Lys	Lys	Gln	Ala	Lys		
	435						440					445					
Lys	Gln	Ala	Lys	Asn	Gln	Arg	Arg	Gln	Gln	Lys	Ile	Gln	Gly	Lys	Val		
	450					455				460							
Leu	His	Leu	Asn	Asp	Ile	Cys	Thr	Ile	Asp	His	Pro	Glu	Asp	Ser	Glu		
465					470					475					480		
Tyr	Glu	Ala	Glu	Met	Ser	Leu	Gln	Gly	Glu	Val	Asn	Ile	Lys	Ser	Asn		
			485					490						495			
His	Ile	Ser	Gln	Glu	Gly	Val	Met	His	Lys	Glu	Tyr	Cys	Val	Asn	Gln		
			500					505					510				
Lys	Asp	Leu	Asn	Gly	Gln	Ala	Lys	Met	Ile	Glu	Ser	Val	Thr	Asp	Asn		
		515					520					525					
Gln	Lys	Ser	Thr	Glu	Glu	Val	Asp	Met	Lys	Asn	Ile	Asn	Met	Asp	Asn		
	530					535						540					
Asp	Leu	Glu	Val	Leu	Thr	Ser	Ser	Pro	Thr	Arg	Asn	Leu	Asn	Gly	Ala		
545					550					555					560		
Tyr	Leu	Thr	Glu	Gly	Ser	Asn	Gly	Glu	Val	Asp	Ile	Ser	Asn	Gly	Phe		
			565					570						575			
Lys	Asn	Leu	Asn	Leu	Asn	Ala	Ala	Leu	His	Pro	Asp	Glu	Ile	Asn	Ile		
			580					585					590				
Glu	Ile	Leu	Asn	Asp	Ser	His	Thr	Pro	Gly	Thr	Lys	Val	Tyr	Glu	Val		
	595						600					605					
Val	Asn	Glu	Asp	Pro	Glu	Thr	Ala	Phe	Cys	Thr	Leu	Ala	Asn	Arg	Glu		
	610					615					620						
Val	Phe	Asn	Thr	Asp	Glu	Cys	Ser	Ile	Gln	His	Cys	Leu	Tyr	Gln	Phe		
625					630					635					640		
Thr	Arg	Asn	Glu	Lys	Leu	Arg	Asp	Ala	Asn	Lys	Leu	Leu	Cys	Glu	Val		
			645					650						655			
Cys	Thr	Arg	Arg	Gln	Cys	Asn	Gly	Pro	Lys	Ala	Asn	Ile	Lys	Gly	Glu		
			660					665					670				
Arg	Lys	His	Val	Tyr	Thr	Asn	Ala	Lys	Lys	Gln	Met	Leu	Ile	Ser	Leu		
	675					680						685					
Ala	Pro	Pro	Val	Leu	Thr	Leu	His	Leu	Lys	Arg	Phe	Gln	Gln	Ala	Gly		
	690					695					700						
Phe	Asn	Leu	Arg	Lys	Val	Asn	Lys	His	Ile	Lys	Phe	Pro	Glu	Ile	Leu		
705					710					715					720		
Asp	Leu	Ala	Pro	Phe	Cys	Thr	Leu	Lys	Cys	Lys	Asn	Val	Ala	Glu	Glu		
			725					730						735			

Asn Thr Arg Val Leu Tyr Ser Leu Tyr Gly Val Val Glu His Ser Gly  
 740 745 750  
 Thr Met Arg Ser Gly His Tyr Thr Ala Tyr Ala Lys Ala Arg Thr Ala  
 755 760 765  
 Asn Ser His Leu Ser Asn Leu Val Leu His Gly Asp Ile Pro Gln Asp  
 770 775 780  
 Phe Glu Met Glu Ser Lys Gly Gln Trp Phe His Ile Ser Asp Thr His  
 785 790 795 800  
 Val Gln Ala Val Pro Thr Thr Lys Val Leu Asn Ser Gln Ala Tyr Leu  
 805 810 815  
 Leu Phe Tyr Glu Arg Ile Leu  
 820

&lt;210&gt; 352

&lt;211&gt; 2903

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 352

```

ctcaattcgt caccaggagg aagacggagc tggctgcccc gcccaaaggc ccatgagggg 60
atgcagttat gggctctgtc gccgtggatt gttattttgt gtcagtaagt aatccataaa 120
gtgccaatcat ggaaagaaa cggacaaagg gaaaaactgt tccaatcgat gattcctctg 180
aaacttttaga acctgtgtgc agacacatta gaaaaggatt ggaacaaggc aatttgaaaa 240
aggcttttagt gaatgtggaa tggaaatatc gccaaactgt taagactgac aataaagtga 300
aagataaagc tgaagaagaa acagaagaaa agccttcagt ttggctgtgt cttaaatgtg 360
gccatcaggg ctgtggcaga aattctcagg agcagcatgc cttgaagcac tatctgacgc 420
caagatctga acctcactgt ctggttctta gtttgacaa ctggagtgtg tgggtgttacg 480
tatgtgataa tgagggtccag tattgtagtt caaacagtt gggcaagtg gttgattatg 540
tcagaaaaca agccagcatt acaactccaa agccagcaga gaaagataat ggaaatattg 600
aacttgaaaa taaaaaatta gaaaaagaga gtaagaatga acaagagaga gaaaagaagg 660
aaaacatggc taaagagaat cctcccatga attctccttg ccaaataacc gtgaaaggac 720
tcagtaattt gggaaacaca tgtttcttca atgcagttat gcagaacttg tcacaaacac 780
cagtgccttag agaactacta aaagaagtga aaatgtctgg aacaattgta aaaattgaac 840
cacctgattt ggcattaaca gaaccattag aataaaacct tgagcctcca ggccctctta 900
ctttagccat gagccagttt cttaatgaga tgcaagagac caaaaagggg gttgtgacac 960
cgaaagaact cttttctcag gtctgtaaaa aagcagtgcg gtttaaggc tatcagcagc 1020
aagacagcca ggagctgctt cgctacttat tggatgggat gagagcagaa gaacaccaa 1080
gagtgagtaa aggaataact aaagcatttg gtaattctac tgaaaagttg gatgaagaac 1140
taaaaaataa agttaaagat tatgagaaga aaaaatcaat gccaaagttt gttgaccgca 1200
tctttggtgg tgaactaact agtatgatca tgtgtgatca atgcagaact gtctccttgg 1260
ttcatgaatc ttctcttgat ttgtccctcc cagtttttaga tgatcagagt ggtaagaaaa 1320
gtgtaaatga taaaaactctg aaaaagacag tggaggatga agatcaagat agtgaggaag 1380
aaaaagataa cgacagttac ataaaagaga gaagtgatat tccttctgga acaagtaagc 1440
acttacagaa aaaagcaaag aaacaagcca aaaagcaagc caagaacca cgaagacaac 1500
aaaaaattca aggaaaagtt cttcatttaa atgatatttg tactattgac catcctgaag 1560
acagtgaata tgaagctgaa atgtcacttc aaggagaagt aaatattaaa tccaaccata 1620
tttcacaaga ggggtgttatg cataaagaat attgtgtcaa ccagaaagat ttgaatggcc 1680
aagcaaaaat gatcgaaagt gtaactgaca atcaaaaatc cacagaggaa gtagatatga 1740
aaaatatcaa catggataat gatctggagg ttttaacatc ttctcccatc aggaatttaa 1800
atggtgccta cctaacggaa gggagcaatg gagaagtgga catttccaat ggtttcaaaa 1860
acctaaattt gaatgctgct cttcatcctg atgaaataaa tatagagatt ctgaatgata 1920
gtcatactcc tggaacaaag gtgtatgagg ttgtaaatga agatccagaa actgctttct 1980
gtactcttgc aaacagggaa gttttcaata ctgatgagtg ttcaatccaa cattgtttat 2040
atcagttcac ccgtaatgag aaacttcgag atgcgaataa actgctttgt gaagtatgca 2100
cacggagaca gtgtaatgga ccaaaggcaa atataaaagg tgaaaggaag catgtttaca 2160
ccaatgccaa aaagcagatg ctaatttctc ttgtcctcc tggttcttact cttcatttaa 2220
agagatttca gcaggctggg tttaacctac gcaaagttaa caaacacata aagtttccgg 2280
aaatcttaga tttggctcct ttttgacccc ttaaatgtaa gaatgttgca gaagaaaata 2340
caagggtagt ctattcctta tatggagttg ttgaacacag tggtagtatg aggtcggggc 2400
attacactgc ctatgcaatg gcaagaacag caaatagatca tctctctaag cttgtttctt 2460

```

```

acggtgatat tccacaagat ttgaaatgg aatcaaaagg gcagtgggtt cacatcagcg 2520
acacacatgt gcaagctgtg cctacaacta aagtactaaa ctcacaagcg tacctcctat 2580
tttatgagag aatactgtaa taatatcaaa agcacttttt ctggaaacac atttatggct 2640
tttataatgg ctgaaataac gataaaaaaa gactaattaa aatcatgttc acttaacatt 2700
aaatacatgc cagaagaaat catgtttatt taaatattga agggaaaaat acctaaaaat 2760
gtacaaaggt tttatattgt catagtgggt tttattcctg ctttgtttct ggaaaggaaa 2820
tcctgaatta cttaagtact ttgtgtttta tatatctggg tgatggatca caacacatca 2880
ataaactgac ttaccctaaa atc
2903

```

&lt;210&gt; 353

&lt;211&gt; 724

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 353

```

Met Ser Phe Pro Lys Ala Pro Leu Lys Arg Phe Asn Asp Pro Ser Gly
1      5      10      15
Cys Ala Pro Ser Pro Gly Ala Tyr Asp Val Lys Thr Leu Glu Val Leu
20     25     30
Lys Gly Pro Val Ser Phe Gln Lys Ser Gln Arg Phe Lys Gln Gln Lys
35     40     45
Glu Ser Lys Gln Asn Leu Asn Val Asp Lys Asp Thr Thr Leu Pro Ala
50     55     60
Ser Ala Arg Lys Val Lys Ser Ser Glu Ser Lys Glu Ser Gln Lys Asn
65     70     75     80
Asp Lys Asp Leu Lys Ile Leu Glu Lys Glu Ile Arg Val Leu Leu Gln
85     90     95
Glu Arg Gly Ala Gln Asp Ser Arg Ile Gln Asp Leu Glu Thr Glu Leu
100    105    110
Glu Lys Met Glu Ala Arg Leu Asn Ala Ala Leu Arg Glu Lys Thr Ser
115    120    125
Leu Ser Ala Asn Asn Ala Thr Leu Glu Lys Gln Leu Ile Glu Leu Thr
130    135    140
Arg Thr Asn Glu Leu Leu Lys Ser Lys Phe Ser Glu Asn Gly Asn Gln
145    150    155    160
Lys Asn Leu Arg Ile Leu Ser Leu Glu Leu Met Lys Leu Arg Asn Lys
165    170    175
Arg Glu Thr Lys Met Arg Gly Met Met Ala Lys Gln Glu Gly Met Glu
180    185    190
Met Lys Leu Gln Val Thr Gln Arg Ser Leu Glu Glu Ser Gln Gly Lys
195    200    205
Ile Ala Gln Leu Glu Gly Lys Leu Val Ser Ile Glu Lys Glu Lys Ile
210    215    220
Asp Glu Lys Ser Glu Thr Glu Lys Leu Leu Glu Tyr Ile Glu Glu Ile
225    230    235    240
Ser Cys Ala Ser Asp Gln Val Glu Lys Tyr Lys Leu Asp Ile Ala Gln
245    250    255
Leu Glu Glu Asn Leu Lys Glu Lys Asn Asp Glu Ile Leu Ser Leu Lys
260    265    270
Gln Ser Leu Glu Glu Asn Ile Val Ile Leu Ser Lys Gln Val Glu Asp
275    280    285
Leu Asn Val Lys Cys Gln Leu Leu Glu Lys Glu Lys Glu Asp His Val
290    295    300
Asn Arg Asn Arg Glu His Asn Glu Asn Leu Asn Ala Glu Met Gln Asn
305    310    315    320
Leu Lys Gln Lys Phe Ile Leu Glu Gln Gln Glu Arg Glu Lys Leu Gln
325    330    335
Gln Lys Glu Leu Gln Ile Asp Ser Leu Leu Gln Gln Glu Lys Glu Leu
340    345    350
Ser Ser Ser Leu His Gln Lys Leu Cys Ser Phe Gln Glu Glu Met Val
355    360    365

```



Lys Glu Lys Asn Leu Phe Glu Glu Glu Leu Lys Gln Thr Leu Asp Glu  
 370 375 380  
 Leu Asp Lys Leu Gln Gln Lys Glu Glu Gln Ala Glu Arg Leu Val Lys  
 385 390 395 400  
 Gln Leu Glu Glu Glu Ala Lys Ser Arg Ala Glu Glu Leu Lys Leu Leu  
 405 410 415  
 Glu Glu Lys Leu Lys Gly Lys Glu Ala Glu Leu Glu Lys Ser Ser Ala  
 420 425 430  
 Ala His Thr Gln Ala Thr Leu Leu Gln Glu Lys Tyr Asp Ser Met  
 435 440 445  
 Val Gln Ser Leu Glu Asp Val Thr Ala Gln Phe Glu Ser Tyr Lys Ala  
 450 455 460  
 Leu Thr Ala Ser Glu Ile Glu Asp Leu Lys Leu Glu Asn Ser Ser Leu  
 465 470 475 480  
 Gln Glu Lys Ala Ala Lys Ala Gly Lys Asn Ala Glu Asp Val Gln His  
 485 490 495  
 Gln Ile Leu Ala Thr Glu Ser Ser Asn Gln Glu Tyr Val Arg Met Leu  
 500 505 510  
 Leu Asp Leu Gln Thr Lys Ser Ala Leu Lys Glu Thr Glu Ile Lys Glu  
 515 520 525  
 Ile Thr Val Ser Phe Leu Gln Lys Ile Thr Asp Leu Gln Asn Gln Leu  
 530 535 540  
 Lys Gln Gln Glu Glu Asp Phe Arg Lys Gln Leu Glu Asp Glu Glu Gly  
 545 550 555 560  
 Arg Lys Ala Glu Lys Glu Asn Thr Thr Ala Glu Leu Thr Glu Glu Ile  
 565 570 575  
 Asn Lys Trp Arg Leu Leu Tyr Glu Glu Leu Tyr Asn Lys Thr Lys Pro  
 580 585 590  
 Phe Gln Leu Gln Leu Asp Ala Phe Glu Val Glu Lys Gln Ala Leu Leu  
 595 600 605  
 Asn Glu His Gly Ala Ala Gln Glu Gln Leu Asn Lys Ile Arg Asp Ser  
 610 615 620  
 Tyr Ala Lys Leu Leu Gly His Gln Asn Leu Lys Gln Lys Ile Lys His  
 625 630 635 640  
 Val Val Lys Leu Lys Asp Glu Asn Ser Gln Leu Lys Ser Glu Val Ser  
 645 650 655  
 Lys Leu Arg Cys Gln Leu Ala Lys Lys Lys Gln Ser Glu Thr Lys Leu  
 660 665 670  
 Gln Glu Glu Leu Asn Lys Val Leu Gly Ile Lys His Phe Asp Pro Ser  
 675 680 685  
 Lys Ala Phe His His Glu Ser Lys Glu Asn Phe Ala Leu Lys Thr Pro  
 690 695 700  
 Leu Lys Glu Gly Asn Thr Asn Cys Tyr Arg Ala Pro Met Glu Cys Gln  
 705 710 715 720  
 Glu Ser Trp Lys

<210> 354  
 <211> 3002  
 <212> DNA  
 <213> Homo sapiens

<400> 354  
 gccagtcacc ttcagtttct ggagctggcc gtcaacatgt cctttcctaa ggcgcccttg 60  
 aaacgattca atgacccttc tggttgtgca ccattctccag gtgcttatga tgttaaaact 120  
 ttagaagtat tgaagaggacc agtatccttt cagaaatcac aaagatttaa acaacaaaaa 180  
 gaatctaaac aaaatcttaa tgttgacaaa gatactacct tgcctgcttc agctagaaaa 240  
 gttaagtctt cggaatcaaa ggaatctcaa aagaatgata aagatttgaa gatattagag 300  
 aaagagattc gtgttcttct acaggaacgt ggtgccagg acagccggat ccaggatctg 360  
 gaaactgagt tggaaaagat ggaagcaagg ctaaagtctg cactaaggga aaaaacatct 420

```

ctctctgcaa ataatgctac actggaaaaa caacttattg aattgaccag gactaatgaa 480
ctactaaaaat ctaagttttc tgaaaaatgt aaccagaaga atttgagaat tctaagcttg 540
gagttgatga aacttagaaa caaaagagaa acaaagatga ggggtatgat ggctaagcaa 600
gaaggcatgg agatgaagct gcaggtcacc caaaggagtc tcgaagagtc tcaagggaaa 660
atagcccaac tggagggaaa acttgtttca atagagaaaag aaaagattga tgaaaaatct 720
gaaacagaaa aactcttggg atacatcgaa gaaattagtt gtgcttcaga tcaagtggaa 780
aaatacaagc tagatattgc ccagttagaa gaaaatttga aagagaagaa tgatgaaatt 840
ttaagcctta agcagtcctc tgaggagaat attgtttat tatctaaaca agtagaagat 900
ctaaatgtga aatgtcagct gcttgaaaaa gaaaaagaag accatgtcaa caggaataga 960
gaacacaacg aaaatctaaa tgcagagatg caaaacttaa aacagaagtt tattcttgaa 1020
caacaggaac gtgaaaagct tcaacaaaaa gaattacaaa ttgattcact tctgcaacaa 1080
gagaaagaat tatcttcgag tcttcacatg aagctctgtt cttttcaaga ggaaatgggt 1140
aaagagaaga atctgtttga ggaagaatta aagcaaacac tggatgagct tgataaatta 1200
cagcaaaagg aggaacaagc tgaaaggctg gtcaagcaat tggaaagagg agcaaaatct 1260
agagctgaag aattaaaact cctagaagaa aagctgaaag ggaaggaggc tgaactggag 1320
aaaagtagtg ctgctcatat ccaggccacc ctgcttttgc aggaaaagta tgacagtatg 1380
gtgcaaagcc ttgaagatgt tactgctcaa tttgaaagct ataaagcggt aacagccagt 1440
gagatagaag atcttaagct ggagaactca tcattacagg aaaaagcggc caaggctggg 1500
aaaaatgcag aggatgttca gcacagatg ttggcaactg agagctcaaa tcaagaatat 1560
gtaaggatgc ttctagatct gcagaccaag tcagcactaa aggaaacaga aattaaagaa 1620
atcacagttt cttttcttca aaaaataact gatttgcaga accaactcaa gcaacaggag 1680
gaagacttta gaaaacagct ggaagatgaa gaaggaagaa aagctgaaaa agaaaatata 1740
acagcagaat taactgaaga aattaacaag tggcgtctcc tctatgaaga actatataat 1800
aaaacaaaac cttttcagct acaactagat gcttttgaag tagaaaaaca ggcattgttg 1860
aatgaacatg gtgcagctca ggaacagcta aataaaaata gagattcata tgctaaatta 1920
ttgggtcatc agaatttgaa acaaaaaatc aagcatgttg tgaagttgaa agatgaaaat 1980
agccaactca aatcggaagt atcaaaactc cgctgtcagc ttgctaaaaa aaaacaaagt 2040
gagacaaaac ttcaagagga attgaataaa gttctaggta tcaaacactt tgatccttca 2100
aaggcttttc atcatgaaag taaagaaaat tttgccctga agaaccattt aaaagaaggc 2160
aatacaaaact gttaccgagc tcctatggag tgtcaagaat catggaagta aacatctgag 2220
aaacctgttg aagattatct ccttcgtctt gttgttattg atgttgctgt tattatattt 2280
gacatgggta ttttataatg ttgtatttaa ttttaactgc caatecttaa atatgtgaaa 2340
ggaacatttt ttaccaaagt gtcttttgac attttatttt ttcttgcaaa tacctcctcc 2400
ctaagtctca cttttatcac ctcatctga accctttcgc tggctttcca gcttagaag 2460
catctcatca acttaaaagt cagtatcata ttattatcct cctgttctga aaccttagtt 2520
tcaagagtct aaaccccaga ttcttcagct tgatcctgga ggtcttttct agtctgagct 2580
tcttttagcta ggctaaaaca ccttggcttg ttattgcctc tactttgatt ctgataatgc 2640
tcacttggtc ctacctatta tccttctact tgtccagttc aaataagaaa taaggacaag 2700
cctaacttca tagaaacctc tctattttta atcagttgtt taataattta caggttctta 2760
ggctccatcc tgtttgtatg aaattataat ctgtgatttg gcctttaagc ctgcattctt 2820
aacaactctc tcagttaatt cttagatata ctaaaaatct gagaaactct acatgtaact 2880
atcttctcag agtttgtcat atactgcttg tcatctgcat gtctactcag catttgatta 2940
acatttgtgt aatatgaaat aaaattacac agtaagtcat ttaaccaaaa aaaaaaaaaa 3000
aa
3002

```

&lt;210&gt; 355

&lt;211&gt; 846

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 355

```

Met Ser Ser Ser His Phe Ala Ser Arg His Arg Lys Asp Ile Ser Thr
1           5           10           15
Glu Met Ile Arg Thr Lys Ile Ala His Arg Lys Ser Leu Ser Gln Lys
20           25           30
Glu Asn Arg His Lys Glu Tyr Glu Arg Asn Arg His Phe Gly Leu Lys
35           40           45
Asp Val Asn Ile Pro Thr Leu Glu Gly Arg Ile Leu Val Glu Leu Asp
50           55           60
Glu Thr Ser Gln Glu Leu Val Pro Glu Lys Thr Asn Val Lys Pro Arg
65           70           75           80

```

Ala Met Lys Thr Ile Leu Gly Asp Gln Arg Lys Gln Met Leu Gln Lys  
 85 90 95  
 Tyr Lys Glu Glu Lys Gln Leu Gln Lys Leu Lys Glu Gln Arg Glu Lys  
 100 105 110  
 Ala Lys Arg Gly Ile Phe Lys Val Gly Arg Tyr Arg Pro Asp Met Pro  
 115 120 125  
 Cys Phe Leu Leu Ser Asn Gln Asn Ala Val Lys Ala Glu Pro Lys Lys  
 130 135 140  
 Ala Ile Pro Ser Ser Val Arg Ile Thr Arg Ser Lys Ala Lys Asp Gln  
 145 150 155 160  
 Met Glu Gln Thr Lys Ile Asp Asn Glu Ser Asp Val Arg Ala Ile Arg  
 165 170 175  
 Pro Gly Pro Arg Gln Thr Ser Glu Lys Lys Val Ser Asp Lys Glu Lys  
 180 185 190  
 Lys Val Val Gln Pro Val Met Pro Thr Ser Leu Arg Met Thr Arg Ser  
 195 200 205  
 Ala Thr Gln Ala Ala Lys Gln Val Pro Arg Thr Val Ser Ser Thr Thr  
 210 215 220  
 Ala Arg Lys Pro Val Thr Arg Ala Ala Asn Glu Asn Glu Pro Glu Gly  
 225 230 235 240  
 Lys Val Pro Ser Lys Gly Arg Pro Ala Lys Asn Val Glu Thr Lys Pro  
 245 250 255  
 Asp Lys Gly Ile Ser Cys Lys Val Asp Ser Glu Glu Asn Thr Leu Asn  
 260 265 270  
 Ser Gln Thr Asn Ala Thr Ser Gly Met Asn Pro Asp Gly Val Leu Ser  
 275 280 285  
 Lys Met Glu Asn Leu Pro Glu Ile Asn Thr Ala Lys Ile Lys Gly Lys  
 290 295 300  
 Asn Ser Phe Ala Pro Lys Asp Phe Met Phe Gln Pro Leu Asp Gly Leu  
 305 310 315 320  
 Lys Thr Tyr Gln Val Thr Pro Met Thr Pro Arg Ser Ala Asn Ala Phe  
 325 330 335  
 Leu Thr Pro Ser Tyr Thr Trp Thr Pro Leu Lys Thr Glu Val Asp Glu  
 340 345 350  
 Ser Gln Ala Thr Lys Glu Ile Leu Ala Gln Lys Cys Lys Thr Tyr Ser  
 355 360 365  
 Thr Lys Thr Ile Gln Gln Asp Ser Asn Lys Leu Pro Cys Pro Leu Gly  
 370 375 380  
 Pro Leu Thr Val Trp His Glu Glu His Val Leu Asn Lys Asn Glu Ala  
 385 390 395 400  
 Thr Thr Lys Asn Leu Asn Gly Leu Pro Ile Lys Glu Val Pro Ser Leu  
 405 410 415  
 Glu Arg Asn Glu Gly Arg Ile Ala Gln Pro His His Gly Val Pro Tyr  
 420 425 430  
 Phe Arg Asn Ile Leu Gln Ser Glu Thr Glu Lys Leu Thr Ser His Cys  
 435 440 445  
 Phe Glu Trp Asp Arg Lys Leu Glu Leu Asp Ile Pro Asp Asp Ala Lys  
 450 455 460  
 Asp Leu Ile Arg Thr Ala Val Gly Gln Thr Arg Leu Leu Met Lys Glu  
 465 470 475 480  
 Arg Phe Lys Gln Phe Glu Gly Leu Val Asp Asp Cys Glu Tyr Lys Arg  
 485 490 495  
 Gly Ile Lys Glu Thr Thr Cys Thr Asp Leu Asp Gly Phe Trp Asp Met  
 500 505 510  
 Val Ser Phe Gln Ile Glu Asp Val Ile His Lys Phe Asn Asn Leu Ile  
 515 520 525  
 Lys Leu Glu Glu Ser Gly Trp Gln Val Asn Asn Asn Met Asn His Asn  
 530 535 540  
 Met Asn Lys Asn Val Phe Arg Lys Lys Val Val Ser Gly Ile Ala Ser  
 545 550 555 560  
 Lys Pro Lys Gln Asp Asp Ala Gly Arg Ile Ala Ala Arg Asn Arg Leu

```
<210> 356
<211> 2979
<212> DNA
<213> Homo sapiens
```

agcaaaccaa	tcgcaagcct	cgttgagtgg	aaggggtggg	atcttcccg	gaagttttgg	60
ttaaagcccc	tccaatcagc	ggctcgggtc	ggcaagtttg	aatttcgtgg	aggctcgggt	120
tgtgaggggt	cctgcttcgg	agtcggcgg	ggtcgtccag	accgagtgtt	ctttactttt	180
tgtttggttg	aggtttcacg	ctagaagggtg	gctcaggatg	tcttcatcac	attttgccag	240
tcgacacag	aaggatataa	gtactgaaat	gattagaact	aaaattgctc	ataggaaatc	300
actgtctcag	aaagaaaaat	gacataagga	atacgaacga	aatagacact	ttggtttgaa	360
agatgtaaac	attccaacct	tggaggtag	aattcttgtt	gaattagatg	agacatctca	420
agagcttgtt	ccagaaaaga	ccaatgttaa	gccaaaggca	atgaaaacta	ttctaggtga	480
tcaacgaaaa	cagatgctcc	aaaaatacaa	agaagaaaag	caacttcaa	atgtgaaaga	540
gcagagagag	aaagctaaac	gaggaatatt	taaagtgggt	cgttatagac	cttatatgcc	600
ttgttttctt	ttatcaaacc	agaatgctgt	gaaagctgag	ccaaaaaagg	ctattccatc	660
ttctgtacgg	attacaaggt	caaaggccaa	agaccaaagt	gagcagacta	agattgataa	720
cgagagtgat	gttcgagcaa	tccgacctgg	tccaagacaa	acttctgaaa	agaaagtgtc	780
agacaaaag	aaaaaagggt	tgcagcctgt	aatgcccacg	tcgttgagaa	tgactcgatc	840
agctactcaa	gcagcaaagc	aggttccag	aacagtctca	tctaccacag	caagaaagcc	900
agtcacaaga	gtcgctaatt	aaaacgaaac	agaaggaaga	gtccaagta	aaggaagacc	960
tgccaaaaat	gtagaaacaa	aacccgacaa	gggtatttct	tgtaaatctg	atagtgaa	1020
aaatactttg	aattcacaaa	ctaattgcaac	aagtggaatg	aatccagatg	gagtccttatc	1080

```

aaaaatggaa aacttacctg agataaatac tgcaaaaata aaaggggaaga attccttcgc 1140
acctaaggat tttatgtttc agccactgga tgggtctgaag acctatcaag taacacctat 1200
gactcccaga agtgccaatg cttttttgac acccagttac acctggactc ctttaaaaac 1260
agaagttgat gagtctcaag caacaaaaga aattttggca caaaaatgta aaacttactc 1320
taccaagaca atacagcaag attcaaataa attgccatgt cctttgggtc ctctaactgt 1380
ttggcatgaa gaacatgttt taaataaaaa tgaagctact actaaaaatt taaatggcct 1440
tccaataaaa gaagtcccat cacttgaaag aaatgaaggt cgaattgctc agccccacca 1500
tgggtgtgcca tatttcagaa atatcctcca gtcagaaact gagaaattaa cttcacattg 1560
cttcgagtgg gacaggaaac ttgaattgga cattccagat gatgctaaag atcttattcg 1620
cacagcagtt ggtcaaaca gactccttat gaaggaaagg tttaaacagt ttgaaggact 1680
ggttgatgat tgtgaatata aacgaggtat aaaggagact acctgtacag atctggatgg 1740
attttgggat atggttagtt ttcagataga agatgtaatc cacaaattca acaatctgat 1800
caaaacttgag gaatctgggt ggcaagtcaa taataatatg aatcataata tgaacaaaaa 1860
tgtctttagg aaaaaagttg tctcaggtat agcaagtaaa ccaaaacagg atgatgctgg 1920
aagaattgca gcgagaaatc gcctagctgc cataaaaaat gcaatgagag agagaattag 1980
gcaggagaagaa tgtgctgaaa cagcagtttc tgtgatacca aaggaagttg ataaaatagt 2040
gttcgatgct ggattttttca gagttgaaag tctgtttaa ttattctcag gactttctgt 2100
ctcttctgaa ggcccttctc aaagacttgg aacacctaag tctgtcaaca aagctgtatc 2160
tcagagtaga aatgagatgg gcattccaca acaaaactaca tcaccagaaa atgccggtcc 2220
tcagaatacg aaaagtgaac atgtgaagaa gactttgttt ttgagtattc ctgaaagcag 2280
gagcagcata gaagatgctc agtgtcctgg attaccagat ttaattgaag aaaaccatgt 2340
tgtaataaag acagacttga aggtggattg tttatccagt gagagaatga gtttgcctct 2400
tcttgctggt ggagtagcag atgatattaa tactaaca aaagaaggaa tttcagatgt 2460
tgtggaagga atggaactga attcttcaat tacatcacag gatgttttga tgagttagccc 2520
tgaaaaaaat acagcttcac aaaatagcat cttagaagaa ggggaaacta aaatttctca 2580
gtcagaacta tttgataata aaagtctcac tactgaatgc caccttcttg attcaccagg 2640
tctaaactgc agtaatccat ttactcagct ggagaggaga catcaagaac atgccagaca 2700
catttctttt ggtggttaacc tgattacttt ttcacctcta caaccaggag aattttgaat 2760
ttaaataata atccaaacat tttccttcac attatcaatg cttatatatt ccttagacta 2820
ttgaaatttt ggagaaaatg tatttggttt cacttctata gcatataatg ttttaattatt 2880
ctgtgttcat caaagtgtat tttagatata ctctttctca agggaagtg ggaattttt 2940
tacattttca acacagaata aaaaatgtac tgtgccttg 2979

```

&lt;210&gt; 357

&lt;211&gt; 191

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 357

```

Met Ala Thr Leu Ile Tyr Val Asp Lys Glu Ile Gly Glu Pro Gly Thr
.1          5          10          15
Arg Val Ala Ala Lys Asp Val Leu Lys Leu Glu Ser Arg Pro Ser Ile
          20          25          30
Lys Ala Leu Asp Gly Ile Ser Gln Val Leu Thr Pro Arg Phe Gly Lys
          35          40          45
Thr Tyr Asp Ala Pro Ser Ala Leu Pro Lys Ala Thr Arg Lys Ala Leu
          50          55          60
Gly Thr Val Asn Arg Ala Thr Glu Lys Ser Val Lys Thr Asn Gly Pro
65          70          75          80
Arg Lys Gln Lys Gln Pro Ser Phe Ser Ala Lys Lys Met Thr Glu Lys
          85          90          95
Thr Val Lys Thr Lys Ser Ser Val Pro Ala Ser Asp Asp Ala Tyr Pro
          100          105          110
Glu Ile Glu Lys Phe Phe Pro Phe Asn Leu Leu Asp Phe Glu Ser Phe
          115          120          125
Asp Leu Pro Glu Glu Arg Gln Ile Ala His Leu Pro Leu Ser Gly Val
          130          135          140
Pro Leu Met Ile Leu Asp Glu Glu Gly Glu Leu Glu Lys Leu Phe Gln
145          150          155          160
Leu Gly Pro Pro Ser Pro Val Lys Met Pro Ser Pro Pro Trp Glu Cys
          165          170          175

```

Asn Leu Phe Ala Val Ser Phe Lys His Ser Val Asp Pro Gly Cys  
 180 185 190

<210> 358  
 <211> 576  
 <212> DNA  
 <213> Homo sapiens

<400> 358  
 atggctactc tgatctacgt tgataaggaa attggagAAC caggcaccCG tgtggctgcc 60  
 aaggatgtgc tgaagctgga gtctagacct tcaatcaaag cattagatgg gatattctca 120  
 gttttaacac cacgttttgg caaaacatac gatgctccat cagccttacc taaagctacc 180  
 agaaaggctt tgggcactgt caacagagct acagaaaagt cagtaaagac caatggaccc 240  
 agaaaacaaa aacagccaag cttttctgcc aaaaagatga ccgagaagac tgttaaaaca 300  
 aaaagtcttg ttcctgcctc agatgacgcc tatccagaaa tagaaaaatt ctttccttcc 360  
 aatcttctag actttgagag ttttgacctg cctgaagagc gccagattgc acacctcccc 420  
 ttgagtggag tgcctctcat gatccttgat gaggaggag agcttgaaaa gctgtttcag 480  
 ctgggcccc cttcacctgt gaaaatgcc tctccaccat gggaatgcaa tctgtttgca 540  
 gtctccttca agcattctgt cgaccctgga tgttga 576

<210> 359  
 <211> 202  
 <212> PRT  
 <213> Homo sapiens

<400> 359  
 Met Ala Thr Leu Ile Tyr Val Asp Lys Glu Asn Glu Glu Pro Gly Ile  
 1 5 10 15  
 Leu Val Ala Thr Lys Asp Gly Leu Lys Leu Gly Ser Gly Pro Ser Ile  
 20 25 30  
 Lys Ala Leu Asp Gly Arg Ser Gln Val Ser Ile Ser Cys Phe Gly Lys  
 35 40 45  
 Thr Phe Asp Ala Pro Thr Ser Leu Pro Lys Ala Thr Arg Lys Ala Leu  
 50 55 60  
 Gly Thr Val Asn Arg Ala Thr Glu Lys Ser Val Lys Thr Asn Gly Pro  
 65 70 75 80  
 Leu Lys Gln Lys Gln Pro Ser Phe Ser Ala Lys Lys Met Thr Glu Lys  
 85 90 95  
 Thr Val Lys Ala Lys Asn Ser Val Pro Ala Ser Asp Asp Gly Tyr Pro  
 100 105 110  
 Glu Ile Glu Lys Leu Phe Pro Phe Asn Pro Leu Gly Phe Glu Ser Phe  
 115 120 125  
 Asp Leu Pro Glu Glu His Gln Ile Ala His Leu Pro Leu Ser Glu Val  
 130 135 140  
 Pro Leu Met Ile Leu Asp Glu Glu Arg Glu Leu Glu Lys Leu Phe Gln  
 145 150 155 160  
 Leu Gly Pro Pro Ser Pro Leu Lys Met Pro Ser Pro Pro Trp Lys Ser  
 165 170 175  
 Asn Leu Leu Gln Ser Pro Leu Ser Ile Leu Leu Thr Leu Asp Val Glu  
 180 185 190  
 Leu Pro Pro Val Cys Cys Asp Ile Asp Ile  
 195 200

<210> 360  
 <211> 609  
 <212> DNA  
 <213> Homo sapiens

<400> 360

```

atggctactc tgatctatgt tgataaggaa aacgaagaac caggcatcct tgtggctaca 60
aaggatgggc tgaagctggg gtctggacct tcaatcaaag ccttagatgg gagatctcaa 120
gtttcaatat catgttttgg caaaacattc gatgctccca catccttacc taaagctacc 180
agaaaggcctt tgggaactgt caacagagct acagaaaagt cagtaaagac caatggaccc 240
ctcaaaaaaa aacagccaag cttttctgcc aaaaagatga ctgagaagac tgttaaagca 300
aaaaactctg ttcctgcctc agatgatggc tatccagaaa tagaaaaatt atttcccttc 360
aatcctctag gcttcgagag ttttgacctg cctgaagagc accagattgc acatctcccc 420
ttgagtgaag tgcctctcat gatacttgat gaggagagag agcttgaaaa gctgtttcag 480
ctgggccccc ctccaccttt gaagatgccc tctccaccat ggaaatccaa tctgttgacg 540
tctcctttaa gcattctggt gaccctggat gttgaattgc cacctgtttg ctgtgacata 600
gatatttaa                                     609

```

&lt;210&gt; 361

&lt;211&gt; 450

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 361

```

Met Asp Ala Arg Arg Val Pro Gln Lys Asp Leu Arg Val Lys Lys Asn
 1          5          10          15
Leu Lys Lys Phe Arg Tyr Val Lys Leu Ile Ser Met Glu Thr Ser Ser
 20          25          30
Ser Ser Asp Asp Ser Cys Asp Ser Phe Ala Ser Asp Asn Phe Ala Asn
 35          40          45
Thr Lys Pro Lys Phe Arg Ser Asp Ile Ser Glu Glu Leu Ala Ser Val
 50          55          60
Phe Tyr Glu Asp Ser Asp Asn Glu Ser Phe Cys Gly Phe Ser Glu Ser
 65          70          75          80
Glu Val Gln Asp Val Leu Asp His Cys Gly Phe Leu Gln Lys Pro Arg
 85          90          95
Pro Asp Val Thr Asn Glu Leu Ala Gly Ile Phe His Ala Asp Ser Asp
100          105          110
Asp Glu Ser Phe Cys Gly Phe Ser Glu Ser Glu Ile Gln Asp Gly Met
115          120          125
Arg Leu Gln Ser Val Arg Glu Gly Cys Arg Thr Arg Ser Gln Cys Arg
130          135          140
His Ser Gly Pro Leu Arg Val Ala Met Lys Phe Pro Ala Arg Ser Thr
145          150          155          160
Arg Gly Ala Thr Asn Lys Lys Ala Glu Ser Arg Gln Pro Ser Glu Asn
165          170          175
Ser Val Thr Asp Ser Asn Ser Asp Ser Glu Asp Glu Ser Gly Met Asn
180          185          190
Phe Leu Glu Lys Arg Ala Leu Asn Ile Lys Gln Asn Lys Ala Met Leu
195          200          205
Ala Lys Leu Met Ser Glu Leu Glu Ser Phe Pro Gly Ser Phe Arg Gly
210          215          220
Arg His Pro Leu Pro Gly Ser Asp Ser Gln Ser Arg Arg Pro Arg Arg
225          230          235          240
Arg Thr Phe Pro Gly Val Ala Ser Arg Arg Asn Pro Glu Arg Arg Ala
245          250          255
Arg Pro Leu Thr Arg Ser Arg Ser Arg Ile Leu Gly Ser Leu Asp Ala
260          265          270
Leu Pro Met Glu Glu Glu Glu Glu Glu Asp Lys Tyr Met Leu Val Arg
275          280          285
Lys Arg Lys Thr Val Asp Gly Tyr Met Asn Glu Asp Asp Leu Pro Arg
290          295          300
Ser Arg Arg Ser Arg Ser Ser Val Thr Leu Pro His Ile Ile Arg Pro
305          310          315          320
Val Glu Glu Ile Thr Glu Glu Glu Leu Glu Asn Val Cys Ser Asn Ser
325          330          335
Arg Glu Lys Ile Tyr Asn Arg Ser Leu Gly Ser Thr Cys His Gln Cys

```

340 345 350  
 Arg Gln Lys Thr Ile Asp Thr Lys Thr Asn Cys Arg Asn Pro Asp Cys  
 355 360 365  
 Trp Gly Val Arg Gly Gln Phe Cys Gly Pro Cys Leu Arg Asn Arg Tyr  
 370 375 380  
 Gly Glu Glu Val Arg Asp Ala Leu Leu Asp Pro Asn Trp His Cys Pro  
 385 390 395 400  
 Pro Cys Arg Gly Ile Cys Asn Cys Ser Phe Cys Arg Gln Arg Asp Gly  
 405 410 415  
 Arg Cys Ala Thr Gly Val Leu Val Tyr Leu Ala Lys Tyr His Gly Phe  
 420 425 430  
 Gly Asn Val His Ala Tyr Leu Lys Ser Leu Lys Gln Glu Phe Glu Met  
 435 440 445  
 Gln Ala  
 450

<210> 362  
 <211> 2824  
 <212> DNA  
 <213> Homo sapiens

<400> 362  
 cccgagcccc gcccctccgg gcccgggtcg gcgcgcccag cctgccagcc gcgctgctgc 60  
 tgctcctcct gctgtgggac cgctgaccgc gcggctgctc cgctctcccc gctccaagcg 120  
 ccgatctggg caccgcgccac cagcatggac gctcgccgcg tgccgcagaa agatctcaga 180  
 gtaaagaaga acttaaagaa attcagatat gtgaagttga tttccatgga aacctcgtca 240  
 tcctctgatg acagttgtga cagctttgct tctgataatt ttgcaaaccac gaaacctaaa 300  
 ttcaggtcag atatcagtga agaactggca agtggttttt atgaggactc tgataatgaa 360  
 tctttctgcg gcttttcaga aagtgtggtg caagatgtat tagaccattg tggattttta 420  
 cagaaaccaa ggagacatgt cactaacgaa ctggccggta tttttcatgc cgactctgac 480  
 gatgaatcat tttgcggttt ctgagagagt tgcaggcact ctggacctct cagggtggcg 600  
 gttcgggaag gctgtaggac ccgcagccag gcaaccaaca aaaaagcaga gtcccgccag 660  
 atgaagtttc cagcgcggag taccagggga tccgattcag aagatgaaag tggaaatgaa 720  
 ccctcagaga attctgtgac tgattccaac tccgattcag aagatgaaag tggaaatgaa 720  
 tttttggaga aaagggcttt aaatataaag caaaacaaag caatgcttgc aaaactcatg 780  
 tctgaattag aaagcttccc tggctcgttc cgtggaagac atccccctcc aggctccgac 840  
 tcacaatcaa ggagaccgcg aaggcgtaca aggtcccga tctcgggtc gagaaacct 900  
 gaacggagag ctgctcctct taccaggtca aggtcccga tctcgggtc ccttgacgct 960  
 ctacccatgg aggaggagga ggaagaggat aagtacatgt tggtagagaa gaggaagacc 1020  
 gtggatggct acatgaatga agatgacctg ccagaaagcc gtcgctccag atcatccgtg 1080  
 acccttccgc atataattcg cccagtggaa gaaattacag aggaggagt ggagaacgct 1140  
 tgcagcaatt ctgagagaa gatataaac cgttccactg gctctacttg tcatcaatgc 1200  
 cgtcagaaga ctattgatac caaaacaaac tgcagaaacc cagactgctg gggcggtcga 1260  
 ggccagttct gtggcccctg ctttcgaaac cgttatggtg aagaggtcag ggatgctctg 1320  
 ctggatccga actggcattg cccgccttgt cgaggaatct gcaactgcag tttctgccg 1380  
 cagcgagatg gacggtgtgc gactggggtc cttgtgtatt tagccaaata tcatggcttt 1440  
 gggaatgtgc atgcctactt gaaaagcctg aaacaggaat ttgaaatgca agcataatat 1500  
 ctggaaaatt tgctgcctgc cttctacttc tcaaactctt cttgtaaaag tttccaattt 1560  
 tttcactgaa acctgagtta aaaatcttga tgatcagcct gtttcataag aaactccaat 1620  
 caagttaatc ttagcagaca tgtgtttctg gagcatcaca gaaggtatat tgctagttag 1680  
 actttgccct cctgcagttt cttctctgct cccaaccccc atctcatagc atccccctct 1740  
 atttccaatg ctctctcca accgcttagt ttctgaattt cttttaaatt acagttttat 1800  
 gaaagcatat tttatttact tgggtgtgaa atagccctca taaaacctaa gcacttgga 1860  
 acacaataat agtattaact aactagatct attgaatttc agagaagagc cttctaactt 1920  
 gtttacacaa aaacgagtat gatttagcat tcatactagt tgaaattttt aatagaatca 1980  
 aggcacaaaa gtcttaaaac catgtggaaa aattaggtaa ttattgcaga ttgatgtctc 2040  
 tcaatcccat gtattgcgct tatgtttaca cgtttgtgca cagttgagac ttaatttctc 2100  
 ctaatttctt ctgcccgaag ggtaagtggc gcgtccagct tacacaatca taattcaaag 2160  
 gttggtgggc aatgtaatac ttaattaaaa taatgatgga agagctatct ggagattatg 2220  
 agtaagctga tttgaatttt cagtataaaa ctttagtata attgtagttt gcaaagttta 2280



```

tttcagttca catgtaaggt attgcaaata aattcttggg caatttttga tggaaacttg 2340
atattaaaaa ctagtctgtg gttcttttgca gtttcttgta aattttataaa ccaggcacaa 2400
ggttcaagtt tagatttttaa gcacttttat aacaatgata agtgcctttt tggagatgta 2460
acttttagca gtttggttaac ctgacatctc tgccagtcta gtttctgggc aggtttcctg 2520
tgtcagtatt cccctcctc tttgcatata tcaagggtatt tggtagaggt ggaatctaag 2580
tgtttgatg tccaatttac ttgcatatgt aaaccattgc tgtgccattc aatgtttgat 2640
gcataattgg accttgaatc gataagtgt aatacagctt ttgatctgta atgcttttat 2700
acaaaagttt attttaataa taaaatgttt gttctaactt gtctgctttt ttaaaaataa 2760
tcttactgta cttaattcta attttttctt catatttaaa taaaaggcca tttccacctt 2820
ttct 2824

```

&lt;210&gt; 363

&lt;211&gt; 371

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 363

```

Met Asp Ala Arg Arg Val Pro Gln Lys Asp Leu Arg Val Lys Lys Asn
 1          5          10          15
Leu Lys Lys Phe Arg Tyr Val Lys Leu Ile Ser Met Glu Thr Ser Ser
 20          25          30
Ser Ser Asp Asp Ser Cys Asp Ser Phe Ala Ser Asp Asn Phe Ala Asn
 35          40          45
Thr Arg Leu Gln Ser Val Arg Glu Gly Cys Arg Thr Arg Ser Gln Cys
 50          55          60
Arg His Ser Gly Pro Leu Arg Val Ala Met Lys Phe Pro Ala Arg Ser
 65          70          75          80
Thr Arg Gly Ala Thr Asn Lys Lys Ala Glu Ser Arg Gln Pro Ser Glu
 85          90          95
Asn Ser Val Thr Asp Ser Asn Ser Asp Ser Glu Asp Glu Ser Gly Met
100          105          110
Asn Phe Leu Glu Lys Arg Ala Leu Asn Ile Lys Gln Asn Lys Ala Met
115          120          125
Leu Ala Lys Leu Met Ser Glu Leu Glu Ser Phe Pro Gly Ser Phe Arg
130          135          140
Gly Arg His Pro Leu Pro Gly Ser Asp Ser Gln Ser Arg Arg Pro Arg
145          150          155          160
Arg Arg Thr Phe Pro Gly Val Ala Ser Arg Arg Asn Pro Glu Arg Arg
165          170          175
Ala Arg Pro Leu Thr Arg Ser Arg Ser Arg Ile Leu Gly Ser Leu Asp
180          185          190
Ala Leu Pro Met Glu Glu Glu Glu Glu Glu Asp Lys Tyr Met Leu Val
195          200          205
Arg Lys Arg Lys Thr Val Asp Gly Tyr Met Asn Glu Asp Asp Leu Pro
210          215          220
Arg Ser Arg Arg Ser Arg Ser Ser Val Thr Leu Pro His Ile Ile Arg
225          230          235          240
Pro Val Glu Glu Ile Thr Glu Glu Glu Leu Glu Asn Val Cys Ser Asn
245          250          255
Ser Arg Glu Lys Ile Tyr Asn Arg Ser Leu Gly Ser Thr Cys His Gln
260          265          270
Cys Arg Gln Lys Thr Ile Asp Thr Lys Thr Asn Cys Arg Asn Pro Asp
275          280          285
Cys Trp Gly Val Arg Gly Gln Phe Cys Gly Pro Cys Leu Arg Asn Arg
290          295          300
Tyr Gly Glu Glu Val Arg Asp Ala Leu Leu Asp Pro Asn Trp His Cys
305          310          315          320
Pro Pro Cys Arg Gly Ile Cys Asn Cys Ser Phe Cys Arg Gln Arg Asp
325          330          335
Gly Arg Cys Ala Thr Gly Val Leu Val Tyr Leu Ala Lys Tyr His Gly
340          345          350

```

Phe Gly Asn Val His Ala Tyr Leu Lys Ser Leu Lys Gln Glu Phe Glu  
 355 360 365  
 Met Gln Ala  
 370

<210> 364  
 <211> 2587  
 <212> DNA  
 <213> Homo sapiens

<400> 364  
 cccgagcccc gccctccg gcccgggtcg gcgcgcccag cctgccagcc gcgtgctgc 60  
 tgctcctcct gctgtgggac cgctgaccgc gcggctgctc cgctctcccc gctccaagcg 120  
 ccgatctggg caccgcccac cagcatggac gctcgccgcg tgccgcagaa agatctcaga 180  
 gtaaagaaga acttaaagaa attcagatat gtgaagtga tttccatgga aacctcgtca 240  
 tcctctgatg acagttgtga cagctttgct tctgataatt ttgcaaacac gaggctgcag 300  
 tcagttcggg aaggctgtag gaccgcagc cagtgcaggc actctggacc tctcaggggtg 360  
 gcgatgaagt ttccagcgcg gactaccagg ggagcaacca acaaaaaagc agagtcccgc 420  
 cagccctcag agaattctgt gactgattcc aactccgatt cagaagatga aagtggatg 480  
 aatttttttg agaaaagggc tttaaatata aagcaaaaca aagcaatgct tgcaaaactc 540  
 atgtctgaat tagaaaagctt ccctggctcg ttccgtggaa gacatccccct cccaggctcc 600  
 gactcacaat caaggagacc gcgaaggcgt acattcccgc gtgttgcttc caggagaaac 660  
 cctgaacgga gagctcgctc tcttaccagg tcaaggctcc ggatcctcgg gtcccttgac 720  
 gctctaccca tggaggagga ggagggaag gataagtaca tgttggtgag aaagaggaag 780  
 accgtggatg gctacatgaa tgaagatgac ctgccagaa gccgtcgctc cagatcatcc 840  
 gtgacccttc cgcataataat tcgcccagtg gaagaaatta cagaggagga gttggagaac 900  
 gtctgcagca attctcgaga gaagatatat aaccgttcac tgggctctac ttgtcatcaa 960  
 tgccgtcaga agactattga taccaaaaca aactgcagaa acccagactg ctggggcggtt 1020  
 cgaggccagt tctgtggccc ctgccttcga aaccgttatg gtgaagaggt cagggatgct 1080  
 ctgctggatc cgaactggca ttgccgcctt tgcagaggaa tctgcaactg cagtttctgc 1140  
 cggcagcgag atggacgggtg tgcgactggg gtccttgggt atttagccaa atatcatggc 1200  
 tttgggaatg tgcattgccta cttgaaaagc ctgaaacagg aatttgaaat gcaagcataa 1260  
 tatctgaaaa atttgctgcc tgccttctac ttctcaaatc tttcttgtaa aagtttccaa 1320  
 ttttttcaat gaaacctgag ttaaaaaatct tgatgatcag cctgtttcat aagaaactcc 1380  
 aatcaagtta atcttagcag acatgtgttt ctggagcatc acagaaggta tattgctagt 1440  
 tacactttgc cctcctgcag tttcttctct gctcccaacc cccatctcat agcatcccc 1500  
 tctatttcca atgtcctctt ccaaccgctt tttcttctgaa tttcttttaa attacagttt 1560  
 tatgaaagca tattttattt acttgggtgtt gaaatagccc tcataaaaacc taagcacttg 1620  
 gaaacacaat aatagtatta actaactaga tctattgaat ttcagagaag agccttctaa 1680  
 cttgtttaca caaaaacgag tatgatttag cattcatact agttgaaatt tttaatagaa 1740  
 tcaaggcaca aaagtcttaa aaccatgtgg aaaaattagg taattattgc agattgatgt 1800  
 ctctcaatcc catgtattgc gcttatgtta caagtgttg tccagttga gacttaattt 1860  
 ctctcaattt cttctgcccg aagggttaagt ggtgcgtcca gcttacacaa tcataattca 1920  
 aagggttggtg ggcaatgtaa tacttaatta aaataatgat ggaagagcta tctggagatt 1980  
 atgagtaagc tgatttgaat ttccagtata aaactttagt ataattgtag ttgcaaaagt 2040  
 ttatttcagt tcacatgtaa ggtattgcaa ataaattctt ggacaatttt gtatggaaac 2100  
 ttgatattaa aaactagtct gtggttcttt gcagtttctt gtaaatttat aaaccaggca 2160  
 caaggttcaa gtttagattt taagcacttt tataacaatg ataagtgcct ttttgagat 2220  
 gtaactttta gcagtttggt aacctgacat ctctgccagt ctagtctctg ggcaggtttc 2280  
 ctgtgtcagt attccccctc ctctttgcat taatcaaggt atttggtaga ggtggaatct 2340  
 aagtgtttgt atgtccaatt tacttgcata tgtaaaccat tgctgtgcca tccaatgttt 2400  
 gatgcataat tggaccttga atcgataagt gtaaatacag cttttgatct gtaatgcttt 2460  
 tatacaaaag tttattttta taataaaatg tttgttctaa cttgtctgct tttttaaaaa 2520  
 taactctact gtacttaatt ctaatttttt cctcatattt aaataaaagg ccatttccac 2580  
 cttttct 2587

<210> 365  
 <211> 709  
 <212> PRT  
 <213> Homo sapiens

&lt;400&gt; 365

```

Met Ser Phe Pro Lys Ala Pro Leu Lys Arg Phe Asn Asp Pro Ser Gly
 1          5          10          15
Cys Ala Pro Ser Pro Gly Ala Tyr Asp Val Lys Thr Leu Glu Val Leu
 20          25          30
Lys Gly Pro Val Ser Phe Gln Lys Ser Gln Arg Phe Lys Gln Gln Lys
 35          40          45
Glu Ser Lys Gln Asn Leu Asn Val Asp Lys Asp Thr Thr Leu Pro Ala
 50          55          60
Ser Ala Arg Lys Val Lys Ser Ser Glu Ser Lys Ile Arg Val Leu Leu
 65          70          75          80
Gln Glu Arg Gly Ala Gln Asp Ser Arg Ile Gln Asp Leu Glu Thr Glu
 85          90          95
Leu Glu Lys Met Glu Ala Arg Leu Asn Ala Ala Leu Arg Glu Lys Thr
100          105          110
Ser Leu Ser Ala Asn Asn Ala Thr Leu Glu Lys Gln Leu Ile Glu Leu
115          120          125
Thr Arg Thr Asn Glu Leu Leu Lys Ser Lys Phe Ser Glu Asn Gly Asn
130          135          140
Gln Lys Asn Leu Arg Ile Leu Ser Leu Glu Leu Met Lys Leu Arg Asn
145          150          155          160
Lys Arg Glu Thr Lys Met Arg Gly Met Met Ala Lys Gln Glu Gly Met
165          170          175
Glu Met Lys Leu Gln Val Thr Gln Arg Ser Leu Glu Glu Ser Gln Gly
180          185          190
Lys Ile Ala Gln Leu Glu Gly Lys Leu Val Ser Ile Glu Lys Glu Lys
195          200          205
Ile Asp Glu Lys Ser Glu Thr Glu Lys Leu Leu Glu Tyr Ile Glu Glu
210          215          220
Ile Ser Cys Ala Ser Asp Gln Val Glu Lys Tyr Lys Leu Asp Ile Ala
225          230          235          240
Gln Leu Glu Glu Asn Leu Lys Glu Lys Asn Asp Glu Ile Leu Ser Leu
245          250          255
Lys Gln Ser Leu Glu Glu Asn Ile Val Ile Leu Ser Lys Gln Val Glu
260          265          270
Asp Leu Asn Val Lys Cys Gln Leu Leu Glu Lys Glu Lys Glu Asp His
275          280          285
Val Asn Arg Asn Arg Glu His Asn Glu Asn Leu Asn Ala Glu Met Gln
290          295          300
Asn Leu Lys Gln Lys Phe Ile Leu Glu Gln Gln Glu Arg Glu Lys Leu
305          310          315          320
Gln Gln Lys Glu Leu Gln Ile Asp Ser Leu Leu Gln Gln Glu Lys Glu
325          330          335
Leu Ser Ser Ser Leu His Gln Lys Leu Cys Ser Phe Gln Glu Glu Met
340          345          350
Val Lys Glu Lys Asn Leu Phe Glu Glu Leu Lys Gln Thr Leu Asp
355          360          365
Glu Leu Asp Lys Leu Gln Gln Lys Glu Glu Gln Ala Glu Arg Leu Val
370          375          380
Lys Gln Leu Glu Glu Glu Ala Lys Ser Arg Ala Glu Glu Leu Lys Leu
385          390          395          400
Leu Glu Glu Lys Leu Lys Gly Lys Glu Ala Glu Leu Glu Lys Ser Ser
405          410          415
Ala Ala His Thr Gln Ala Thr Leu Leu Gln Glu Lys Tyr Asp Ser
420          425          430
Met Val Gln Ser Leu Glu Asp Val Thr Ala Gln Phe Glu Ser Tyr Lys
435          440          445
Ala Leu Thr Ala Ser Glu Ile Glu Asp Leu Lys Leu Glu Asn Ser Ser
450          455          460
Leu Gln Glu Lys Ala Ala Lys Ala Gly Lys Asn Ala Glu Asp Val Gln

```

465                      470                      475                      480  
 His Gln Ile Leu Ala Thr Glu Ser Ser Asn Gln Glu Tyr Val Arg Met  
                                  485                      490                      495  
 Leu Leu Asp Leu Gln Thr Lys Ser Ala Leu Lys Glu Thr Glu Ile Lys  
                                  500                      505                      510  
 Glu Ile Thr Val Ser Phe Leu Gln Lys Ile Thr Asp Leu Gln Asn Gln  
                                  515                      520                      525  
 Leu Lys Gln Gln Glu Glu Asp Phe Arg Lys Gln Leu Glu Asp Glu Glu  
                                  530                      535                      540  
 Gly Arg Lys Ala Glu Lys Glu Asn Thr Thr Ala Glu Leu Thr Glu Glu  
 545                      550                      555                      560  
 Ile Asn Lys Trp Arg Leu Leu Tyr Glu Glu Leu Tyr Asn Lys Thr Lys  
                                  565                      570                      575  
 Pro Phe Gln Leu Gln Leu Asp Ala Phe Glu Val Glu Lys Gln Ala Leu  
                                  580                      585                      590  
 Leu Asn Glu His Gly Ala Ala Gln Glu Gln Leu Asn Lys Ile Arg Asp  
                                  595                      600                      605  
 Ser Tyr Ala Lys Leu Leu Gly His Gln Asn Leu Lys Gln Lys Ile Lys  
 610                      615                      620  
 His Val Val Lys Leu Lys Asp Glu Asn Ser Gln Leu Lys Ser Glu Val  
 625                      630                      635                      640  
 Ser Lys Leu Arg Cys Gln Leu Ala Lys Lys Lys Gln Ser Glu Thr Lys  
                                  645                      650                      655  
 Leu Gln Glu Glu Leu Asn Lys Val Leu Gly Ile Lys His Phe Asp Pro  
                                  660                      665                      670  
 Ser Lys Ala Phe His His Glu Ser Lys Glu Asn Phe Ala Leu Lys Thr  
                                  675                      680                      685  
 Pro Leu Lys Glu Gly Asn Thr Asn Cys Tyr Arg Ala Pro Met Glu Cys  
 690                      695                      700  
 Gln Glu Ser Trp Lys  
 705

&lt;210&gt; 366

&lt;211&gt; 2957

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 366

```

gccagtcacc ttcagtttct ggagctggcc gtcaacatgt cctttcctaa ggcgcccttg 60
aaacgattca atgacccttc tggttgtgca ccatctccag gtgcttatga tgttaaaact 120
ttagaagtat tgaaaggacc agtatccttt cagaaatcac aaagatttaa acaacaaaaa 180
gaatctaaac aaaatcttaa tgttgacaaa gatactacct tgcctgcttc agctagaaaa 240
gttaagtctt cggaatcaaa gattcgtggt cttctacagg aacgtggtgc ccaggacagc 300
cggtaccagg atctggaaac tgagttggaa aagatggaa gcaaggctaaa tgctgcacta 360
agggaaaaaa catctctctc tgcaaataat gctacactgg aaaaacaact tattgaattg 420
accaggacta atgaactact aaaatctaag ttttctgaaa atggtaacca gaagaatttg 480
agaattctaa gcttgagggt gatgaaactt agaaacaaaa gagaaacaaa gatgaggggt 540
atgatggcta agcaagaagg catggagatg aagctgcagg tcacccaaag gactctcgaa 600
gagtctcaag ggaaaatagc ccaactggag ggaaaacttg tttcaataga gaaagaaaag 660
attgatgaaa aatctgaaac agaaaaactc ttggaatata tcgaagaaat tagttgtgct 720
tcagatcaag tggaaaaata caagctagat attgcccagt tagaagaaaa ttgaaagag 780
aagaatgatg aaattttaag ccttaagcag tctcttgagg agaattattg tatattatct 840
aaacaagtag aagatctaaa tgtgaaatgt cagctgcttg aaaaagaaaa agaagaccat 900
gtcaacagga atagagaaca caacgaaaat ctaaattgcag agatgcaaaa cttaaaacag 960
aagtttattc ttgaacaaca ggaacgtgaa aagcttcaac aaaaagaatt acaaattgat 1020
tcacttctgc aacaagagaa agaattatct tcgagtcctc atcagaagct ctgttctttt 1080
caagaggaaa tggttaaaga gaagaatctg ttgaggaag aattaaagca aacactggat 1140
gagcttgata aattacagca aaaggaggaa caagctgaaa ggctggtcaa gcaattggaa 1200
gaggaagcaa aatctagagc tgaagaatta aaactcctag aagaaaagct gaaagggag 1260
gaggctgaac tggagaaaaa tagtgctgct catacccagg ccaccctgct tttgcaggaa 1320

```

```

aagtatgaca gtatggtgca aagccttgaa gatgttactg ctcaatttga aagctataaa 1380
gcgttaacag ccagtgagat agaagatcctt aagctggaga actcatcatt acaggaaaaa 1440
gcggccaaagg ctgggaaaaaa tgcagaggat gttcagcatc agattttggc aactgagagc 1500
tcaaatacaag aatatgtaag gatgcttcta gatctgcaga ccaagtcagc actaaaggaa 1560
acagaaatta aagaaatcac agtttctttt cttcaaaaaa taactgattt gcagaaccaa 1620
ctcaagcaac aggaggaaga ctttagaaaa cagctggaag atgaagaagg aagaaaaagc 1680
gaaaaagaaa atacaacagc agaattaact gaagaaatta acaagtggcg tctcctctat 1740
gaagaactat ataataaaac aaaacctttt cagctacaac tagatgcttt tgaagtagaa 1800
aaacaggcat tgttgaaatga acatggtgca gctcaggaac agctaaataa aataagagat 1860
tcatatgcta aattattggg tcatcagaat ttgaaacaaa aaatcaagca tgttgtagag 1920
ttgaaagatg aaaatagcca actcaaactg gaagtatcaa aactccgctg tcagcttgct 1980
aaaaaaaaac aaagtgcagc aaaacttcaa gaggaattga ataaagttct aggtatcaaa 2040
cactttgatc cttcaaaggc ttttcatcat gaaagtaaag aaaattttgc cctgaagacc 2100
ccattaaaag aaggcaatac aaactgttac cgagctccta tggagtgtca agaatacatg 2160
aagtaaacat ctgagaaacc tgttgaagat tatttcattc gtcttgttgt tattgatgtt 2220
gctgttatta tatttgacat gggattttta taatgttgta ttttaatttta actgccaatc 2280
cttaaatatg tgaaagggaac attttttacc aaagtgtctt ttgacatttt attttttctt 2340
gcaaatacct cctccctaata gctcaccttt atcacctcat tctgaaccct ttcgctggct 2400
ttccagctta gaatgcatct catcaactta aaagtcagta tcataattatt atcctcctgt 2460
tctgaaacct tagtttcaag agtctaaacc ccagattcct cagcttgatc ctggagggtc 2520
tttctagtct gagcttcttt agctaggcta aaacaccttg gcttggtatt gcctctactt 2580
tgattctgat aatgctcact tggctcctacc tattatcctt ctacttgctc agttcaaata 2640
agaaataagg acaagcctaa cttcatagaa acctctctat ttttaatcag ttgtttaata 2700
atttacaggt tcttaggctc catcctgttt gtatgaaatt ataactctgt gattggcctt 2760
taagcctgca ttcttaacaa actcttcagt taattcttag atacactaaa aatctgagaa 2820
actctacatg taactatttc ttcagagttt gtcataact gcttgatc tgcagtgcta 2880
ctcagcattt gattaacatt tgtgtaatat gaaataaaat tacacagtaa gtcatttaac 2940
caaaaaaaaa aaaaaaa 2957

```

&lt;210&gt; 367

&lt;211&gt; 79

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 367

```

Met Ser His Lys Gln Ile Tyr Tyr Ser Asp Lys Tyr Asp Asp Glu Glu
  1             5             10             15
Phe Glu Tyr Arg His Val Met Leu Pro Lys Asp Ile Ala Lys Leu Val
      20             25             30
Pro Lys Thr His Leu Met Ser Glu Ser Glu Trp Arg Asn Leu Gly Val
      35             40             45
Gln Gln Ser Gln Gly Trp Val His Tyr Met Ile His Glu Pro Glu Pro
      50             55             60
His Ile Leu Leu Phe Arg Arg Pro Leu Pro Lys Lys Pro Lys Lys
      65             70             75

```

&lt;210&gt; 368

&lt;211&gt; 717

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 368

```

agagcgatca tgtcgcaaaa acaaattttac tattcggaca aatacgacga cgaggagttt 60
gagtatcgac atgtcatgct gcccaaggac atagccaagc tggctccctaa aacccatctg 120
atgtctgaat ctgaatggag gaatcttggc gttcagcaga gtcagggatg ggtccattat 180
atgatccatg aaccagaacc tcacatcttg ctgttccggc gccactacc caagaaacca 240
aagaaatgaa gctggcaagc tacttttcag cctcaagctt tacacagctg tccttacttc 300
ctaactctt tctgataaca ttattatgtt gccttcttgt ttctcacttt gatattttaa 360
agatgttcaa tacactgttt gaatgtgctg gtaactgctt tgcttcttga gtagagccac 420
caccaccata gccagccag atgagtgtctc tgtggaccca cagcctaagc tgagtgtgac 480

```

cccagaagcc acgatgtgct ctgtatccag aacacacttg gcagatggag gaagcatctg 540  
 agtttgagac catggctgtt acagggatca tgtaaacttg ctgtttttgt tttttctgcc 600  
 ggggtgtgta tgtgtgggtga cttgcggatt tatgttttcag tgtactggaa actttccatt 660  
 ttattcaaga aatctgttca tgtaaaaagc cttgattaaa gaggaagttt ttataat 717

<210> 369

<211> 191

<212> PRT

<213> Homo sapiens

<400> 369

Met Ala Thr Leu Ile Tyr Val Asp Lys Glu Ile Gly Glu Pro Gly Thr  
 1 5 10 15  
 Arg Val Ala Ala Lys Asp Val Leu Lys Leu Glu Ser Arg Pro Ser Ile  
 20 25 30  
 Lys Ala Leu Asp Gly Ile Ser Gln Val Leu Thr Pro Arg Phe Gly Lys  
 35 40 45  
 Thr Tyr Asp Ala Pro Ser Ala Leu Pro Lys Ala Thr Arg Lys Ala Leu  
 50 55 60  
 Gly Thr Val Asn Arg Ala Thr Glu Lys Ser Val Lys Thr Asn Gly Pro  
 65 70 75 80  
 Arg Lys Gln Lys Gln Pro Ser Phe Ser Ala Lys Lys Met Thr Glu Lys  
 85 90 95  
 Thr Val Lys Thr Lys Ser Ser Val Pro Ala Ser Asp Asp Ala Tyr Pro  
 100 105 110  
 Glu Ile Glu Lys Phe Phe Pro Phe Asn Leu Leu Asp Phe Glu Ser Phe  
 115 120 125  
 Asp Leu Pro Glu Glu Arg Gln Ile Ala His Leu Pro Leu Ser Gly Val  
 130 135 140  
 Pro Leu Met Ile Leu Asp Glu Glu Gly Glu Leu Glu Lys Leu Phe Gln  
 145 150 155 160  
 Leu Gly Pro Pro Ser Pro Val Lys Met Pro Ser Pro Pro Trp Glu Cys  
 165 170 175  
 Asn Leu Phe Ala Val Ser Phe Lys His Ser Val Asp Pro Gly Cys  
 180 185 190

<210> 370

<211> 576

<212> DNA

<213> Homo sapiens

<400> 370

atggctactc tgatctacgt tgataaggaa attggagaac caggcaccgc tgtggctgcc 60  
 aaggatgtgc tgaagctgga gtctagacct tcaatcaaag cattagatgg gatattctcaa 120  
 gttttaacac cacgttttgg caaaacatac gatgctccat cagccttacc taaagctacc 180  
 agaaaaggctt tgggcaactgt caacagagct acagaaaagt cagttaaagac caatggaccc 240  
 agaaaacaaa aacagccaag cttttctgcc aaaaagatga ccgagaagac tgtaaaaaca 300  
 aaaagttctg ttctgcctc agatgacgcc tatccagaaa tagaaaaatt ctttccttc 360  
 aatcttctag actttgagag ttttgacctg cctgaagagc gccagattgc acacctcccc 420  
 ttgagtggag tgcctctcat gatccttgat gagggaggag agcttgaaaa gctgtttcag 480  
 ctgggcccc cttcacctgt gaaaatgccc tctccaccat gggaatgcaa tctgtttgca 540  
 gtctccttca agcattctgt cgaccctgga tgttga 576

<210> 371

<211> 212

<212> PRT

<213> Homo sapiens

<400> 371

Met Lys Pro Pro Ser Ser Ile Gln Thr Ser Glu Phe Asp Ser Ser Asp

1	5	10	15
Glu Glu Pro Ile Glu Asp Glu Gln Thr Pro Ile His Ile Ser Trp Leu			
	20	25	30
Ser Leu Ser Arg Val Asn Cys Ser Gln Phe Leu Gly Leu Cys Ala Leu			
	35	40	45
Pro Gly Cys Lys Phe Lys Asp Val Arg Arg Asn Val Gln Lys Asp Thr			
	50	55	60
Glu Glu Leu Lys Ser Cys Gly Ile Gln Asp Ile Phe Val Phe Cys Thr			
65	70	75	80
Arg Gly Glu Leu Ser Lys Tyr Arg Val Pro Asn Leu Leu Asp Leu Tyr			
	85	90	95
Gln Gln Cys Gly Ile Ile Thr His His His Pro Ile Ala Asp Gly Gly			
	100	105	110
Thr Pro Asp Ile Ala Ser Cys Cys Glu Ile Met Glu Glu Leu Thr Thr			
	115	120	125
Cys Leu Lys Asn Tyr Arg Lys Thr Leu Ile His Cys Tyr Gly Gly Leu			
	130	135	140
Gly Arg Ser Cys Leu Val Ala Ala Cys Leu Leu Leu Tyr Leu Ser Asp			
145	150	155	160
Thr Ile Ser Pro Glu Gln Ala Ile Asp Ser Leu Arg Asp Leu Arg Gly			
	165	170	175
Ser Gly Ala Ile Gln Thr Ile Lys Gln Tyr Asn Tyr Leu His Glu Phe			
	180	185	190
Arg Asp Lys Leu Ala Ala His Leu Ser Ser Arg Asp Ser Gln Ser Arg			
	195	200	205
Ser Val Ser Arg			
210			

<210> 372  
 <211> 836  
 <212> DNA  
 <213> Homo sapiens

<400> 372  
 ccggtgagtc gccggcgctg cagagggagg cggcactggt ctcgacgtgg ggcggccagc 60  
 gatgaagccg cccagttcaa tacaaacaag tgagtttgac tcatcagatg aagagcctat 120  
 tgaagatgaa cagactccaa ttcataatc atggctatct ttgtcacgag tgaattgttc 180  
 tcagtttctc ggtttatgtg ctcttccagg ttgtaaattt aaagatgtta gaagaaatgt 240  
 ccaaaaagat acagaagaac taaagagctg tggatatacaa gacatatattg ttttctgcac 300  
 cagaggggaa ctgtcaaaat atagagtccc aaaccttctg gatctctacc agcaatgtgg 360  
 aattatcacc catcatcatc caatcgaga tggagggact cctgacatag ccagctgctg 420  
 tgaaataatg gaagagctta caacctgcct taaaaattac cgaaaaacct taatacactg 480  
 ctatggagga cttgggagat cttgtcttgt agctgcttgt ctctactat acctgtctga 540  
 cacaatatca ccagagcaag ccatagacag cctgcgagac ctaagaggat ccggggcaat 600  
 acagaccatc aagcaatata attatcttca tgagtttcgg gacaaattag ctgcacatct 660  
 atcatcaaga gattcacaaat caagatctgt atcaagataa aggaattcaa atagcatata 720  
 tatgaccatg tctgaaatgt cagttctcta gcataatttg tattgaaatg aaaccaccag 780  
 tgttatcaac ttgaatgtaa atgtacatgt gcagatattc ctaaagtttt attgac 836

<210> 373  
 <211> 1085  
 <212> PRT  
 <213> Homo sapiens

<400> 373  
 Met Asp Thr Pro Glu Asn Val Leu Gln Met Leu Glu Ala His Met Gln  
 1 5 10 15  
 Ser Tyr Lys Gly Asn Asp Pro Leu Gly Glu Trp Glu Arg Tyr Ile Gln  
 20 25 30  
 Trp Val Glu Glu Asn Phe Pro Glu Asn Lys Glu Tyr Leu Ile Thr Leu

35 40 45  
 Leu Glu His Leu Met Lys Glu Phe Leu Asp Lys Lys Lys Tyr His Asn  
 50 55 60  
 Asp Pro Arg Phe Ile Ser Tyr Cys Leu Lys Phe Ala Glu Tyr Asn Ser  
 65 70 75 80  
 Asp Leu His Gln Phe Phe Glu Phe Leu Tyr Asn His Gly Ile Gly Thr  
 85 90 95  
 Leu Ser Ser Pro Leu Tyr Ile Ala Trp Ala Gly His Leu Glu Ala Gln  
 100 105 110  
 Gly Glu Leu Gln His Ala Ser Ala Val Leu Gln Arg Gly Ile Gln Asn  
 115 120 125  
 Gln Ala Glu Pro Arg Glu Phe Leu Gln Gln Gln Tyr Arg Leu Phe Gln  
 130 135 140  
 Thr Arg Leu Thr Glu Thr His Leu Pro Ala Gln Ala Arg Thr Ser Glu  
 145 150 155 160  
 Pro Leu His Asn Val Gln Val Leu Asn Gln Met Ile Thr Ser Lys Ser  
 165 170 175  
 Asn Pro Gly Asn Asn Met Ala Cys Ile Ser Lys Asn Gln Gly Ser Glu  
 180 185 190  
 Leu Ser Gly Val Ile Ser Ser Ala Cys Asp Lys Glu Ser Asn Met Glu  
 195 200 205  
 Arg Arg Val Ile Thr Ile Ser Lys Ser Glu Tyr Ser Val His Ser Ser  
 210 215 220  
 Leu Ala Ser Lys Val Asp Val Glu Gln Val Val Met Tyr Cys Lys Glu  
 225 230 235 240  
 Lys Leu Ile Arg Gly Glu Ser Glu Phe Ser Phe Glu Glu Leu Arg Ala  
 245 250 255  
 Gln Lys Tyr Asn Gln Arg Arg Lys His Glu Gln Trp Val Asn Glu Asp  
 260 265 270  
 Arg His Tyr Met Lys Arg Lys Glu Ala Asn Ala Phe Glu Glu Gln Leu  
 275 280 285  
 Leu Lys Gln Lys Met Asp Glu Leu His Lys Lys Leu His Gln Val Val  
 290 295 300  
 Glu Thr Ser His Glu Asp Leu Pro Ala Ser Gln Glu Arg Ser Glu Val  
 305 310 315 320  
 Asn Pro Ala Arg Met Gly Pro Ser Val Gly Ser Gln Gln Glu Leu Arg  
 325 330 335  
 Ala Pro Cys Leu Pro Val Thr Tyr Gln Gln Thr Pro Val Asn Met Glu  
 340 345 350  
 Lys Asn Pro Arg Glu Ala Pro Pro Val Val Pro Pro Leu Ala Asn Ala  
 355 360 365  
 Ile Ser Ala Ala Leu Val Ser Pro Ala Thr Ser Gln Ser Ile Ala Pro  
 370 375 380  
 Pro Val Pro Leu Lys Ala Gln Thr Val Thr Asp Ser Met Phe Ala Val  
 385 390 395 400  
 Ala Ser Lys Asp Ala Gly Cys Val Asn Lys Ser Thr His Glu Phe Lys  
 405 410 415  
 Pro Gln Ser Gly Ala Glu Ile Lys Glu Gly Cys Glu Thr His Lys Val  
 420 425 430  
 Ala Asn Thr Ser Ser Phe His Thr Thr Pro Asn Thr Ser Leu Gly Met  
 435 440 445  
 Val Gln Ala Thr Pro Ser Lys Val Gln Pro Ser Pro Thr Val His Thr  
 450 455 460  
 Lys Glu Ala Leu Gly Phe Ile Met Asn Met Phe Gln Ala Pro Thr Leu  
 465 470 475 480  
 Pro Asp Ile Ser Asp Asp Lys Asp Glu Trp Gln Ser Leu Asp Gln Asn  
 485 490 495  
 Glu Asp Ala Phe Glu Ala Gln Phe Gln Lys Asn Val Arg Ser Ser Gly  
 500 505 510  
 Ala Trp Gly Val Asn Lys Ile Ile Ser Ser Leu Ser Ser Ala Phe His  
 515 520 525



Val	Phe	Glu	Asp	Gly	Asn	Lys	Glu	Asn	Tyr	Gly	Leu	Pro	Gln	Pro	Lys
530						535					540				
Asn	Lys	Pro	Thr	Gly	Ala	Arg	Thr	Phe	Gly	Glu	Arg	Ser	Val	Ser	Arg
545					550					555					560
Leu	Pro	Ser	Lys	Pro	Lys	Glu	Glu	Val	Pro	His	Ala	Glu	Glu	Phe	Leu
				565					570					575	
Asp	Asp	Ser	Thr	Val	Trp	Gly	Ile	Arg	Cys	Asn	Lys	Thr	Leu	Ala	Pro
			580					585					590		
Ser	Pro	Lys	Ser	Pro	Gly	Asp	Phe	Thr	Ser	Ala	Ala	Gln	Leu	Ala	Ser
		595					600					605			
Thr	Pro	Phe	His	Lys	Leu	Pro	Val	Glu	Ser	Val	His	Ile	Leu	Glu	Asp
	610					615					620				
Lys	Glu	Asn	Val	Val	Ala	Lys	Gln	Cys	Thr	Gln	Ala	Thr	Leu	Asp	Ser
625					630					635					640
Cys	Glu	Glu	Asn	Met	Val	Val	Pro	Ser	Arg	Asp	Gly	Lys	Phe	Ser	Pro
				645					650					655	
Ile	Gln	Glu	Lys	Ser	Pro	Lys	Gln	Ala	Leu	Ser	Ser	His	Met	Tyr	Ser
			660					665					670		
Ala	Ser	Leu	Leu	Arg	Leu	Ser	Gln	Pro	Ala	Ala	Gly	Gly	Val	Leu	Thr
		675					680					685			
Cys	Glu	Ala	Glu	Leu	Gly	Val	Glu	Ala	Cys	Arg	Leu	Thr	Asp	Thr	Asp
	690					695					700				
Ala	Ala	Ile	Ala	Glu	Asp	Pro	Pro	Asp	Ala	Ile	Ala	Gly	Leu	Gln	Ala
705					710					715					720
Glu	Trp	Met	Gln	Met	Ser	Ser	Leu	Gly	Thr	Val	Asp	Ala	Pro	Asn	Phe
				725					730					735	
Ile	Val	Gly	Asn	Pro	Trp	Asp	Asp	Lys	Leu	Ile	Phe	Lys	Leu	Leu	Ser
			740					745					750		
Gly	Leu	Ser	Lys	Pro	Val	Ser	Ser	Tyr	Pro	Asn	Thr	Phe	Glu	Trp	Gln
		755					760					765			
Cys	Lys	Leu	Pro	Ala	Ile	Lys	Pro	Lys	Thr	Glu	Phe	Gln	Leu	Gly	Ser
	770					775					780				
Lys	Leu	Val	Tyr	Val	His	His	Leu	Leu	Gly	Glu	Gly	Ala	Phe	Ala	Gln
785					790					795					800
Val	Tyr	Glu	Ala	Thr	Gln	Gly	Asp	Leu	Asn	Asp	Ala	Lys	Asn	Lys	Gln
				805					810					815	
Lys	Phe	Val	Leu	Lys	Val	Gln	Lys	Pro	Ala	Asn	Pro	Trp	Glu	Phe	Tyr
				820				825					830		
Ile	Gly	Thr	Gln	Leu	Met	Glu	Arg	Leu	Lys	Pro	Ser	Met	Gln	His	Met
		835					840					845			
Phe	Met	Lys	Phe	Tyr	Ser	Ala	His	Leu	Phe	Gln	Asn	Gly	Ser	Val	Leu
	850					855					860				
Val	Gly	Glu	Leu	Tyr	Ser	Tyr	Gly	Thr	Leu	Leu	Asn	Ala	Ile	Asn	Leu
865					870				875						880
Tyr	Lys	Asn	Thr	Pro	Glu	Lys	Val	Met	Pro	Gln	Gly	Leu	Val	Ile	Ser
				885					890					895	
Phe	Ala	Met	Arg	Met	Leu	Tyr	Met	Ile	Glu	Gln	Val	His	Asp	Cys	Glu
			900					905					910		
Ile	Ile	His	Gly	Asp	Ile	Lys	Pro	Asp	Asn	Phe	Ile	Leu	Gly	Asn	Gly
		915					920					925			
Phe	Leu	Glu	Gln	Asp	Asp	Glu	Asp	Asp	Leu	Ser	Ala	Gly	Leu	Ala	Leu
	930					935					940				
Ile	Asp	Leu	Gly	Gln	Ser	Ile	Asp	Met	Lys	Leu	Phe	Pro	Lys	Gly	Thr
945					950					955					960
Ile	Phe	Thr	Ala	Lys	Cys	Glu	Thr	Ser	Gly	Phe	Gln	Cys	Val	Glu	Met
				965					970					975	
Leu	Ser	Asn	Lys	Pro	Trp	Asn	Tyr	Gln	Ile	Asp	Tyr	Phe	Gly	Val	Ala
			980					985					990		
Ala	Thr	Val	Tyr	Cys	Met	Leu	Phe	Gly	Thr	Tyr	Met	Lys	Val	Lys	Asn
		995					1000					1005			
Glu	Gly	Gly	Glu	Cys	Lys	Pro	Glu	Gly	Leu	Phe	Arg	Arg	Leu	Pro	His

1010  
 Leu Asp Met Trp Asn Glu Phe Phe His Val Met Leu Asn Ile Pro Asp  
 1025  
 Cys His His Leu Pro Ser Leu Asp Leu Leu Arg Gln Lys Leu Lys Lys  
 1030  
 1040  
 1045  
 Val Phe Gln Gln His Tyr Thr Asn Lys Ile Arg Ala Leu Arg Asn Arg  
 1050  
 1055  
 1060  
 1065  
 1070  
 Leu Ile Val Leu Leu Leu Glu Cys Lys Arg Ser Arg Lys  
 1075  
 1080  
 1085

<210> 374  
 <211> 3446  
 <212> DNA  
 <213> Homo sapiens

<400> 374  
 ttctagtttg cggttcaggt ttgccgctgc cggccagcgt cctctggcca tggacacccc 60  
 ggaaaatgtc cttcagatgc ttgaagccca catgcagagc tacaagggca atgaccctct 120  
 tgggtgaatgg gaaagataca tacagtgggt agaagagaat ttctctgaga ataaagaata 180  
 cttgataact ttactagaac atttaatgaa ggaattttta gataagaaga aataccacaa 240  
 tgacccaaga ttcatcaggt attgtttaaa atttgctgag tacaacagtg acctccatca 300  
 attttttgag ttctgttaca accatgggat tggaaacctg tcatcccctc tgtacattgc 360  
 ctgggcgggg catctggaag cccaaggaga gctgcagcat gccagtgtctg tccttcagag 420  
 aggaattcaa aaccaggctg aaccagagaga ttctctgcaa caacaatata ggttattttca 480  
 gacacgcctc actgaaaccc atttgccagc tcaagctaga acctcagaac ctctgcataa 540  
 tgttcaggtt ttaaatcaaa tgataacatc aaaatcaaat ccaggaaata acatggcctg 600  
 catttctaag aatcagggtt cagagctttc tggagtata tcttcagctt gtgataaaga 660  
 gtcaaatatg gaacgaagag tgatcacgat ttctaaatca gaattattctg tgactcatc 720  
 ttggcatcc aaagttgatg ttgagcaggt tggtatgtat tgcaaggaga agcttattcg 780  
 tggggaatca gaattttcct ttgaagaatt gagagcccag aaatacaatc aacggagaaa 840  
 gcatgagcaa tgggtaaatg aagacagaca ttatatgaaa aggaaagaag caaatgcttt 900  
 tgaagaacag ctattaaaac agaaaatgga tgaacttcat aagaagttgc atcaggtggt 960  
 ggagacatcc catgaggatc tgcccgttc ccaggaaagg tccgaggtta atccagcacg 1020  
 tatggggcca agtgtaggct cccagcagga actgagagcg ccatgtcttc cagtaacctc 1080  
 tcagcagaca ccagtgaaca tggaaaagaa cccaagagag gcacctcctg ttgttcctcc 1140  
 tttggcaaat gctatttctg cagctttggt gtccccagcc accagccaga gcattgctcc 1200  
 tctgttctct ttgaaagccc agacagtaac agactccatg tttgcagtgg ccagcaaaga 1260  
 tgctggatgt gtgaataaga gtactcatga attcaagcca cagagtggag cagagatcaa 1320  
 agaaggggtg gaaacacata aggttgccaa cacaagttct tttcacacaa ctccaaacac 1380  
 atcactggga atggttcagg caacgccatc caaagtgcag ccatcaccca ccgtgcacac 1440  
 aaaagaagca ttaggtttca tcatgaatat gtttcaggct cctacacttc ctgatatttc 1500  
 tgatgacaaa gatgaatggc aatctctaga tcaaaatgaa gatgcatttg aagcccagtt 1560  
 tcaaaaaaat gtaaggtcat ctggggcttg gggagtcaat aagatcatct cttctttgtc 1620  
 atctgctttt catgtgtttg aagatggaaa caaagaaaaa tatggattac cacagcctaa 1680  
 aaataaacc cagaggacca ggacctttgg agaacgctct gtcagcagac ttcttcaaa 1740  
 accaaaggag gaagtgcctc atgctgaaga gtttttggat gactcaactg tatggggtat 1800  
 tcgctgcaac aaaaccctgg caccagctcc taagagccca ggagacttca catctgctgc 1860  
 acaacttgcg tctacacct tccacaagct tccagtggag tcagtgcaca ttttagaaga 1920  
 taaagaaaaa gtggtagcaa aacagtgtac ccaggcgact ttggattctt gtgaggaaaa 1980  
 catggtggtg ccttcaaggg atggaaaatt cagtccaatt caagagaaaa gcccaaaaca 2040  
 ggccttgctg tctcacatgt attcagcatc cttacttctg ctgagccagc ctgctgcagg 2100  
 tggggtactt acctgtgagg cagagtggg cgttgaggct tgcagactca cagacactga 2160  
 cgctgccatt gcagaagatc caccagatgc tattgtgagg aaacttcatt gttgggaacc catgggatga 2220  
 gatgatttca cttgggactg ttgatgctcc ttctaaacca gtgagttcct atccaaatac 2280  
 taagctgatt ttcaaacttt tatctgggct ttccagccat caagcccaag actgaatttc aattgggttc 2340  
 ttttgaatgg caatgtaaac accttcttgg agaaggagcc tttgcccagg tgtacgaagc 2400  
 taagctggtc tatgtccatc accttcttgg agaaggagcc tttgcccagg tgtacgaagc 2460  
 taccagggga gatctgaatg atgctaaaaa taaacagaaa tttgttttaa aggtccaaaa 2520  
 gcctgccaac ccctgggaat tctacattgg gaccagttg atggaaagac taaagccatc 2580  
 tatgcagcac atgtttatga agttctattc tgcccactta ttccagaatg gcagtgtatt 2640

```

agtaggagag ctctacagct atggaacatt attaaatgcc attaacctct ataaaaatac 2700
ccctgaaaaa gtgatgcctc aaggctctgt catctctttt gctatgagaa tgctttacat 2760
gattgagcaa gtgcatgact gtgaaatcat tcatggagac attaaaccag acaatttcat 2820
acttgaaac ggatttttgg aacaggatga tgaagatgat ttatctgctg gcttggcact 2880
gattgacctg ggtcagagta tagatatgaa actttttcca aaaggaacta tattcacagc 2940
aaagtgtgaa acatctgggt ttcagtgtgt tgagatgctc agcaacaaac catggaacta 3000
ccagatcgat tactttgggg ttgctgcaac agtatattgc atgctctttg gcacttacat 3060
gaaagtgaaa aatgaaggag gagagtgtaa gcctgaaggc ctttttagaa ggcttcctca 3120
tttgatatg tggaatgaat tttttcatgt tatgttgaat attccagatt gtcacatctc 3180
tccatctttg gatttggtta ggcaaaagct gaagaaagta tttcaacaac actatactaa 3240
caagattagg gccctacgta ataggcta atgtactgctc ttagaatgta agcggttcacg 3300
aaaataaaat ttggatatag acagtcctta aaaatcacac tgtaaataatg aatctgctca 3360
ctttaaacct gttttttttt cattttattgt ttatgtaaat gtttggttaa aataaatccc 3420
atggaatatt tccatgtaaa aaaaaa 3446

```

&lt;210&gt; 375

&lt;211&gt; 724

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 375

```

Met Ser Phe Pro Lys Ala Pro Leu Lys Arg Phe Asn Asp Pro Ser Gly
 1           5           10           15
Cys Ala Pro Ser Pro Gly Ala Tyr Asp Val Lys Thr Leu Glu Val Leu
          20           25           30
Lys Gly Pro Val Ser Phe Gln Lys Ser Gln Arg Phe Lys Gln Gln Lys
          35           40           45
Glu Ser Lys Gln Asn Leu Asn Val Asp Lys Asp Thr Thr Leu Pro Ala
          50           55           60
Ser Ala Arg Lys Val Lys Ser Ser Glu Ser Lys Glu Ser Gln Lys Asn
          65           70           75           80
Asp Lys Asp Leu Lys Ile Leu Glu Lys Glu Ile Arg Val Leu Leu Gln
          85           90           95
Glu Arg Gly Ala Gln Asp Ser Arg Ile Gln Asp Leu Glu Thr Glu Leu
          100          105          110
Glu Lys Met Glu Ala Arg Leu Asn Ala Ala Leu Arg Glu Lys Thr Ser
          115          120          125
Leu Ser Ala Asn Asn Ala Thr Leu Glu Lys Gln Leu Ile Glu Leu Thr
          130          135          140
Arg Thr Asn Glu Leu Leu Lys Ser Lys Phe Ser Glu Asn Gly Asn Gln
          145          150          155          160
Lys Asn Leu Arg Ile Leu Ser Leu Glu Leu Met Lys Leu Arg Asn Lys
          165          170          175
Arg Glu Thr Lys Met Arg Gly Met Met Ala Lys Gln Glu Gly Met Glu
          180          185          190
Met Lys Leu Gln Val Thr Gln Arg Ser Leu Glu Glu Ser Gln Gly Lys
          195          200          205
Ile Ala Gln Leu Glu Gly Lys Leu Val Ser Ile Glu Lys Glu Lys Ile
          210          215          220
Asp Glu Lys Ser Glu Thr Glu Lys Leu Leu Glu Tyr Ile Glu Glu Ile
          225          230          235          240
Ser Cys Ala Ser Asp Gln Val Glu Lys Tyr Lys Leu Asp Ile Ala Gln
          245          250          255
Leu Glu Glu Asn Leu Lys Glu Lys Asn Asp Glu Ile Leu Ser Leu Lys
          260          265          270
Gln Ser Leu Glu Glu Asn Ile Val Ile Leu Ser Lys Gln Val Glu Asp
          275          280          285
Leu Asn Val Lys Cys Gln Leu Leu Glu Lys Glu Lys Glu Asp His Val
          290          295          300
Asn Arg Asn Arg Glu His Asn Glu Asn Leu Asn Ala Glu Met Gln Asn
          305          310          315          320

```

Leu Lys Gln Lys Phe Ile Leu Glu Gln Gln Glu Arg Glu Lys Leu Gln  
 325 330 335  
 Gln Lys Glu Leu Gln Ile Asp Ser Leu Leu Gln Gln Glu Lys Glu Leu  
 340 345 350  
 Ser Ser Ser Leu His Gln Lys Leu Cys Ser Phe Gln Glu Glu Met Val  
 355 360 365  
 Lys Glu Lys Asn Leu Phe Glu Glu Glu Leu Lys Gln Thr Leu Asp Glu  
 370 375 380  
 Leu Asp Lys Leu Gln Gln Lys Glu Glu Gln Ala Glu Arg Leu Val Lys  
 385 390 395 400  
 Gln Leu Glu Glu Glu Ala Lys Ser Arg Ala Glu Glu Leu Lys Leu Leu  
 405 410 415  
 Glu Glu Lys Leu Lys Gly Lys Glu Ala Glu Leu Glu Lys Ser Ser Ala  
 420 425 430  
 Ala His Thr Gln Ala Thr Leu Leu Leu Gln Glu Lys Tyr Asp Ser Met  
 435 440 445  
 Val Gln Ser Leu Glu Asp Val Thr Ala Gln Phe Glu Ser Tyr Lys Ala  
 450 455 460  
 Leu Thr Ala Ser Glu Ile Glu Asp Leu Lys Leu Glu Asn Ser Ser Leu  
 465 470 475 480  
 Gln Glu Lys Ala Ala Lys Ala Gly Lys Asn Ala Glu Asp Val Gln His  
 485 490 495  
 Gln Ile Leu Ala Thr Glu Ser Ser Asn Gln Glu Tyr Val Arg Met Leu  
 500 505 510  
 Leu Asp Leu Gln Thr Lys Ser Ala Leu Lys Glu Thr Glu Ile Lys Glu  
 515 520 525  
 Ile Thr Val Ser Phe Leu Gln Lys Ile Thr Asp Leu Gln Asn Gln Leu  
 530 535 540  
 Lys Gln Gln Glu Glu Asp Phe Arg Lys Gln Leu Glu Asp Glu Glu Gly  
 545 550 555 560  
 Arg Lys Ala Glu Lys Glu Asn Thr Thr Ala Glu Leu Thr Glu Glu Ile  
 565 570 575  
 Asn Lys Trp Arg Leu Leu Tyr Glu Glu Leu Tyr Asn Lys Thr Lys Pro  
 580 585 590  
 Phe Gln Leu Gln Leu Asp Ala Phe Glu Val Glu Lys Gln Ala Leu Leu  
 595 600 605  
 Asn Glu His Gly Ala Ala Gln Glu Gln Leu Asn Lys Ile Arg Asp Ser  
 610 615 620  
 Tyr Ala Lys Leu Leu Gly His Gln Asn Leu Lys Gln Lys Ile Lys His  
 625 630 635 640  
 Val Val Lys Leu Lys Asp Glu Asn Ser Gln Leu Lys Ser Glu Val Ser  
 645 650 655  
 Lys Leu Arg Cys Gln Leu Ala Lys Lys Lys Gln Ser Glu Thr Lys Leu  
 660 665 670  
 Gln Glu Glu Leu Asn Lys Val Leu Gly Ile Lys His Phe Asp Pro Ser  
 675 680 685  
 Lys Ala Phe His His Glu Ser Lys Glu Asn Phe Ala Leu Lys Thr Pro  
 690 695 700  
 Leu Lys Glu Gly Asn Thr Asn Cys Tyr Arg Ala Pro Met Glu Cys Gln  
 705 710 715 720  
 Glu Ser Trp Lys

<210> 376  
 <211> 3002  
 <212> DNA  
 <213> Homo sapiens

<400> 376  
 gccagtcacc tttagtttct ggagctggcc gtcaacatgt cctttcctaa ggccgaccttg 60

```

aaacgattca atgacccttc tggttgtgca ccatctccag gtgcttatga tgttaaaact 120
ttagaagtat tgaaaggacc agtatccttt cagaaatcac aaagatttaa acaacaaaaa 180
gaatctaaac aaaatcttaa tgttgacaaa gatactacct tgcctgcttc agctagaaaa 240
gttaagtctt cggaatcaaa ggaatctcaa aagaatgata aagatttgaa gatattagag 300
aaagagattc gtgttcttct acaggaacgt ggtgcccgag acagccggat ccaggatctg 360
gaaactgagt tggaaaagat ggaagcaagg ctaaagtctg cactaaggga aaaaacatct 420
ctctctgcaa ataagtctac actggaaaaa caacttattg aattgaccag gactaatgaa 480
ctactaaaaa ctaagttttc tgaaaatggt aaccagaaga atttgagaat tctaagcttg 540
gagttgatga aacttagaaa caaaagagaa acaaagatga ggggtatgat ggctaagcaa 600
gaaggcatgg agatgaagct gcaggtcacc caaaggagtc tcgaagagtc tcaagggaaa 660
atagcccaac tggagggaaa acttggttca atagagaaag aaaagattga tgaaaaatct 720
gaaacagaaa aactcttgga atacatcgaa gaaattagtt gtgcttcaga tcaagtggaa 780
aaatacaagc tagatattgc ccagttagaa gaaaatttga aagagaagaa tgatgaaatt 840
ttaagcctta agcagtctct tgaggagaat attgtttat tctctaaaca agtagaagat 900
ctaaatgta aatgtcagct gcttgaaaaa gaaaaagaag accatgtcaa caggaataga 960
gaacacaacg ctgctcataa tgcagagatg caaaacttaa aacagaagtt tattcttgaa 1020
caacaggaac gtgaaaagct tcaacaaaaa gaattacaaa ttgattcact tctgcaacaa 1080
gagaaagaat tatcttcgag tcttcacag aagctctgtt cttttcaaga ggaaatggtt 1140
aaagagaaga atctgtttga ggaagaatta aagcaaacac tggatgagct tgataaatta 1200
cagcaaaagg aggaacaagc tgaaaggctg gtcaagcaat tggaaaggga agcaaaatct 1260
agagctgaag aattaaact cctagaagaa aagctgaaag ggaaggaggc tgaactggag 1320
aaaagtagtg ctgctcatal ccaggccacc ctgcttttgc aggaaaagta tgacagtatg 1380
gtgcaaaagg ttgaagatgt tactgctcaa tttgaaagct ataaagcgtt aacagccagt 1440
gagatagaag atcttaagct ggagaactca tcattacagg aaaaagcggc caaggctggg 1500
aaaaatgcag aggatgttca gcatcagatt ttggcaactg agagctcaaa tcaagaatat 1560
gtaaggatgc ttctagatct gcagaccaag tcagcactaa aggaaacaga aattaaagaa 1620
atcacagttt cttttcttca aaaaataact gatttgcaga accaactcaa gcaacaggag 1680
gaagacttta gaaaacagct ggaagatgaa gaaggaagaa aagctgaaaa agaaaaataca 1740
acagcagaat taactcgaga aattaacaag tggcgtctcc tctatgaaga actatataat 1800
aaaacaaaac cttttcagct acaactagat gcttttgaag tagaaaaaca ggcattgttg 1860
aatgaacatg gtgcagctca ggaacagcta aataaaaata gagattcata tgctaaatta 1920
ttgggtcatc agaatttgaa acaaaaaatc aagcatgttg tgaagttgaa agatgaaaat 1980
agccaactca aatcggaagt atcaaaactc cgctgtcagc ttgctaaaaa aaaacaaagt 2040
gagacaaaac ttcaagagga attgaataaa gttctaggta tcaaacactt tgatccttca 2100
aaggcttttt atcatgaaag taaagaaaat tttgcctga agacccattt aaaagaaggc 2160
aatcaaaact gttacggagc tcctatggag tgtcaagaat catggaagta aacatctgag 2220
aaacctgttg aagattattt cattcgtctt gttgttattg atgttgctgt tattatattt 2280
gacatgggta ttttataatg ttgtatttaa ttttaactgc caatccttaa atatgtgaaa 2340
ggaacatttt ttaccaaaagt gtcttttgac attttatttt ttcttgcaaa tacctcctcc 2400
ctaattgctca cctttatcac ctcatcttga accctttcgc tggctttcca gcttagaatg 2460
catctcatca acttaaaaagt cagtatcata ttattatcct cctgttctga aaccttagtt 2520
tcaagagctc aaaccccaga ttcttcagct tgatcctgga ggtcttttct agtctgagct 2580
tctttagcta ggctaaaaa ccttggttg ttattgcctc tactttgatt ctgataatgc 2640
tcaacttggt ctacctatta tccttctact tgtccagttc aaataagaaa taaggacaag 2700
cctaacttca tagaaacctc tctattttta atcagttgtt taataattta caggttctta 2760
ggctccatcc tgtttgtatg aaattataat ctgtggattg gcctttaagc ctgcattctt 2820
aacaactctt tcagtttaatt cttagataca ctaaaaatct gagaaactct acatgtaact 2880
atctcttcag agtttgtcat atactgcttg tcatctgcat gtctactcag catttgatta 2940
acatttgtgt aatatgaaat aaaattacac agtaagtcat ttaacaaaaa aaaaaaaaaa 3000
aa

```

<210> 377  
 <211> 246  
 <212> PRT  
 <213> Homo sapiens

<400> 377  
 Met Asn Asp Trp Met Pro Ile Ala Lys Glu Tyr Asp Pro Leu Lys Ala  
 1 5 10 15  
 Gly Ser Ile Asp Gly Thr Asp Glu Asp Pro His Asp Arg Ala Val Trp  
 20 25 30

Arg Ala Met Leu Ala Arg Tyr Val Pro Asn Lys Gly Val Ile Gly Asp  
 35 40 45  
 Pro Leu Leu Thr Leu Phe Val Ala Arg Leu Asn Leu Gln Thr Lys Glu  
 50 55 60  
 Asp Lys Leu Lys Glu Val Phe Ser Arg Tyr Gly Asp Ile Arg Arg Leu  
 65 70 75 80  
 Arg Leu Val Arg Asp Leu Val Thr Gly Phe Ser Lys Gly Tyr Ala Phe  
 85 90 95  
 Ile Glu Tyr Lys Glu Glu Arg Ala Val Ile Lys Ala Tyr Arg Asp Ala  
 100 105 110  
 Asp Gly Leu Val Ile Asp Gln His Glu Ile Phe Val Asp Tyr Glu Leu  
 115 120 125  
 Glu Arg Thr Leu Lys Gly Trp Ile Pro Arg Arg Leu Gly Gly Gly Leu  
 130 135 140  
 Gly Gly Lys Lys Glu Ser Gly Gln Leu Arg Phe Gly Gly Arg Asp Arg  
 145 150 155 160  
 Pro Phe Arg Lys Pro Ile Asn Leu Pro Val Val Lys Asn Asp Leu Tyr  
 165 170 175  
 Arg Glu Gly Lys Arg Glu Arg Arg Glu Arg Ser Arg Ser Arg Glu Arg  
 180 185 190  
 His Trp Asp Ser Arg Thr Arg Asp Arg His Asp Arg Gly Arg Glu  
 195 200 205  
 Lys Arg Trp Gln Glu Arg Glu Pro Thr Arg Val Trp Pro Asp Asn Asp  
 210 215 220  
 Trp Glu Arg Glu Arg Asp Phe Arg Asp Asp Arg Ile Lys Gly Arg Glu  
 225 230 235 240  
 Lys Lys Glu Arg Gly Lys  
 245

<210> 378  
 <211> 1509  
 <212> DNA  
 <213> Homo sapiens

<400> 378  
 ggcgaaatctg ctgctaccaa tgtaaagggtc aggccgaggc cggcgcgagg aatctgctgt 60  
 cgctgcagc tgctcgctg tctccgctcg aaggagagccc aagctttgca gagagacagg 120  
 ggtcgccca ggctggagag cggtggcgcg atctcggtc actgcagcgt cgacatttcg 180  
 ggtcaagcga tctccagcc tcggcctctc aaagtgtgtg tattacaggg gtgagccagc 240  
 gcgtgcctgg ccaaaaattt tctaaatttg tataataatt tataattgta atgcattttt 300  
 atagacacca tatgatctaa tcttcacaaa aacctagtga agtgacattt agctacattt 360  
 cacaataaga atcctgaagc tcaaaattta ctgacctcaa ataattcccag cactttggga 420  
 ggctgaggca ggcggtgat ctgacatcag gagtttgaga ccagcttggc caacatggtg 480  
 aaatcctgtc tgtactaaaa atgcaaaaat tagctgggag tggtggtgtg tgtctgtaat 540  
 ccagctact cggcctccca aagtgtgtgg attacaggcg tgagccaccg cgtctggcct 600  
 cagccaaggt ttttaagtaa catatttcag cattggctct acagcgttgc agaacatgaa 660  
 cgattggatg cccatcgcca aggagtatga tccactcaaa gcgggagcga ttgatggcac 720  
 cgatgaagac ccacacgacc gcgcggtctg gagggcaatg ctggcacgat atgtccccc 780  
 caaaggtgtc ataggagatc ccctcctcac cctgtttgtg gccagactaa acttgcagac 840  
 caaggaggac aaattaaagg aagtcttttc ccgctatggt gacatccggc ggcttcggct 900  
 ggtaaggagc ttggtcacag gtttttcaaa gggctacgcc ttcacgaat acaaggagga 960  
 gcgtgcctgt atcaaagctt accgagatgc tgatggcctg gttattgacc agcatgagat 1020  
 atttgtggac tacgagctgg aaaggactct caaagggttg atccctcggc gacttggagg 1080  
 cggctcttggg ggaaaaaagg agtctgggca actgagattt gggggacggg accggccttt 1140  
 tcgaaaacct attaaacttg cagttgttaa aaacgacctc tatagagagg gaaaacggga 1200  
 aaggcgggag cgatctcgat cccgagaaaag aactggggac tcgaggacaa gggatcgaga 1260  
 ccatgacagg ggccgggaga agagatggca agaaagagag ccgaccaggg tgtggcccga 1320  
 caatgactgg gagagagaga gggacttcag agatgacagg atcaagggga gggagaagaa 1380  
 ggaaagaggc aagtagaggc ccaacagcag aaccccaaag tgaagttaca gtggaaatga 1440  
 gtggaggggg attgtctttc aacgcagcgt gagtctaatt gttgaataaa acttactgat 1500

gatcaaaaa

1509

&lt;210&gt; 379

&lt;211&gt; 246

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 379

```

Met Asn Asp Trp Met Pro Ile Ala Lys Glu Tyr Asp Pro Leu Lys Ala
 1      5      10      15
Gly Ser Ile Asp Gly Thr Asp Glu Asp Pro His Asp Arg Ala Val Trp
      20      25      30
Arg Ala Met Leu Ala Arg Tyr Val Pro Asn Lys Gly Val Ile Gly Asp
      35      40      45
Pro Leu Leu Thr Leu Phe Val Ala Arg Leu Asn Leu Gln Thr Lys Glu
      50      55      60
Asp Lys Leu Lys Glu Val Phe Ser Arg Tyr Gly Asp Ile Arg Arg Leu
      65      70      75      80
Arg Leu Val Arg Asp Leu Val Thr Gly Phe Ser Lys Gly Tyr Ala Phe
      85      90      95
Ile Glu Tyr Lys Glu Glu Arg Ala Val Ile Lys Ala Tyr Arg Asp Ala
      100     105     110
Asp Gly Leu Val Ile Asp Gln His Glu Ile Phe Val Asp Tyr Glu Leu
      115     120     125
Glu Arg Thr Leu Lys Gly Trp Ile Pro Arg Arg Leu Gly Gly Gly Leu
      130     135     140
Gly Gly Lys Lys Glu Ser Gly Gln Leu Arg Phe Gly Gly Arg Asp Arg
      145     150     155     160
Pro Phe Arg Lys Pro Ile Asn Leu Pro Val Val Lys Asn Asp Leu Tyr
      165     170     175
Arg Glu Gly Lys Arg Glu Arg Arg Glu Arg Ser Arg Ser Arg Glu Arg
      180     185     190
His Trp Asp Ser Arg Thr Arg Asp Arg Asp His Asp Arg Gly Arg Glu
      195     200     205
Lys Arg Trp Gln Glu Arg Glu Pro Thr Arg Val Trp Pro Asp Asn Asp
      210     215     220
Trp Glu Arg Glu Arg Asp Phe Arg Asp Asp Arg Ile Lys Gly Arg Glu
      225     230     235     240
Lys Lys Glu Arg Gly Lys
      245

```

&lt;210&gt; 380

&lt;211&gt; 967

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 380

```

ggcgaatctg ctgctaccaa tgtaaaggct aggccgaggc cggcgcgagg aatctgctgt 60
cgctgcagc tgctcgctg tctccgtcgg aaggagagccc aagctttgca gagaacatga 120
acgattggat gcccatcgcc aaggagtatg atccactcaa agcgggcagc attgatggca 180
ccgatgaaga cccacacgac cgcgcggtct ggagggcaat gctggcacga tatgtcccca 240
acaaagggtg cataggagat cccctcctca ccctgtttgt ggccagacta aacttgca 300
ccaaggagga caaattaaag gaagtctttt cccgctatgg tgacatccgg cggttcggc 360
tggtcaggga cttggtcaca ggtttttcaa agggctacgc cttcatcgaa tacaaggagg 420
agcgtgccgt gatcaaagct taccgagatg ctgatggcct ggttattgac cagcatgaga 480
tatttgtgga ctacgagctg gaaaggactc tcaaagggtg gatccctcgg cgacttgagg 540
gcggtcttgg gggaaaaaag gagtctgggc aactgagatt tgggggacgg gaccggcctt 600
ttcgaaaacc tattaacttg ccagttgtta aaaacgacct ctatagagag ggaaaacggg 660
aaaggcggga gcgatctcga tcccagaaaa gacactggga ctcgaggaca agggatcgag 720
accatgacag gggccgggag aagagatggc aagaaagaga gccgaccagg gtgtggcccc 780

```

acaatgactg ggagagagag agggacttca gagatgacag gatcaagggg agggagaaga 840  
 aggaaagagg caagtagagg cccaacagca gaaccccaa gtgaagttac agtggaaatg 900  
 agtggagggg gattgtcttt caacgcagcg tgagtcta ggttgaataa aacttactga 960  
 tgatcaa 967

<210> 381

<211> 226

<212> PRT

<213> Homo sapiens

<400> 381

Met Ser Glu Thr Ala Pro Ala Glu Thr Ala Thr Pro Ala Pro Val Glu  
 1 5 10 15  
 Lys Ser Pro Ala Lys Lys Lys Ala Thr Lys Lys Ala Ala Gly Ala Gly  
 20 25 30  
 Ala Ala Lys Arg Lys Ala Thr Gly Pro Pro Val Ser Glu Leu Ile Thr  
 35 40 45  
 Lys Ala Val Ala Ala Ser Lys Glu Arg Asn Gly Leu Ser Leu Ala Ala  
 50 55 60  
 Leu Lys Lys Ala Leu Ala Ala Gly Gly Tyr Asp Val Glu Lys Asn Asn  
 65 70 75 80  
 Ser Arg Ile Lys Leu Gly Leu Lys Ser Leu Val Ser Lys Gly Thr Leu  
 85 90 95  
 Val Gln Thr Lys Gly Thr Gly Ala Ser Gly Ser Phe Lys Leu Asn Lys  
 100 105 110  
 Lys Ala Ala Ser Gly Glu Ala Lys Pro Lys Ala Lys Lys Ala Gly Ala  
 115 120 125  
 Ala Lys Ala Lys Lys Pro Ala Gly Ala Thr Pro Lys Lys Ala Lys Lys  
 130 135 140  
 Ala Ala Gly Ala Lys Lys Ala Val Lys Lys Thr Pro Lys Lys Ala Lys  
 145 150 155 160  
 Lys Pro Ala Ala Ala Gly Val Lys Lys Val Ala Lys Ser Pro Lys Lys  
 165 170 175  
 Ala Lys Ala Ala Ala Lys Pro Lys Lys Ala Thr Lys Ser Pro Ala Lys  
 180 185 190  
 Pro Lys Ala Val Lys Pro Lys Ala Ala Lys Pro Lys Ala Ala Lys Pro  
 195 200 205  
 Lys Ala Ala Lys Pro Lys Ala Ala Lys Lys Lys Ala Ala Ala Lys  
 210 215 220  
 Lys Lys  
 225

<210> 382

<211> 790

<212> DNA

<213> Homo sapiens

<400> 382

aacctgctct ttagatttct agcttattct cttctagcag tttcttgcca ccatgtcgga 60  
 aaccgctcct gccgagacag ccaccccagc gccggtggag aaatccccgg ctaagaagaa 120  
 ggcaactaag aaggctgccg gcgcccgcgc tgctaagcgc aaagcgacgg ggccccaggt 180  
 ctacagagctg atcaccaagg ctgtggctgc ttctaaggag cgcaatggcc tttctttggc 240  
 agcccttaag aaggccttag cggccggtgg ctacgacgtg gagaagaata acagccgcgt 300  
 taagctgggc ctcaagagct tggtagacaa gggcaccctg gtgcagacca agggcactgg 360  
 tgcttctggc tcctttaaac tcaacaagaa ggcggcctcc ggggaagcca agcccaaagc 420  
 caagaaggca ggcgcgcta aagctaagaa gcccgcgagg gccacgccta agaaggccaa 480  
 gaaggctgca gggcgcaaaa aggcagtgaa gaagactccg aagaaggcga agaagcccgc 540  
 ggcggctggc gtcaaaaagg tggcgaagag ccctaagaag gccaaaggcc ctgccaaacc 600  
 gaaaaaggca accaagagtc ctgccaaagg caaggcagtt aagccgaagg cggcaaagcc 660  
 caaagccgct aagcccaaag cagcaaaaacc taaagctgca aaggccaaga aggcggctgc 720



caaaaagaag taggaagctg gcgtgtgaaa accgcaacaa agccccaag gctcttttca 780  
gagccaccca 790

<210> 383  
<211> 202  
<212> PRT  
<213> Homo sapiens

<400> 383  
Met Ala Thr Leu Ile Tyr Val Asp Lys Glu Asn Glu Glu Pro Gly Ile  
1 5 10 15  
Leu Val Ala Thr Lys Asp Gly Leu Lys Leu Gly Ser Gly Pro Ser Ile  
20 25 30  
Lys Ala Leu Asp Gly Arg Ser Gln Val Ser Ile Ser Cys Phe Gly Lys  
35 40 45  
Thr Phe Asp Ala Pro Thr Ser Leu Pro Lys Ala Thr Arg Lys Ala Leu  
50 55 60  
Gly Thr Val Asn Arg Ala Thr Glu Lys Ser Val Lys Thr Asn Gly Pro  
65 70 75 80  
Leu Lys Gln Lys Gln Pro Ser Phe Ser Ala Lys Lys Met Thr Glu Lys  
85 90 95  
Thr Val Lys Ala Lys Asn Ser Val Pro Ala Ser Asp Asp Gly Tyr Pro  
100 105 110  
Glu Ile Glu Lys Leu Phe Pro Phe Asn Pro Leu Gly Phe Glu Ser Phe  
115 120 125  
Asp Leu Pro Glu Glu His Gln Ile Ala His Leu Pro Leu Ser Glu Val  
130 135 140  
Pro Leu Met Ile Leu Asp Glu Glu Arg Glu Leu Glu Lys Leu Phe Gln  
145 150 155 160  
Leu Gly Pro Pro Ser Pro Leu Lys Met Pro Ser Pro Pro Trp Lys Ser  
165 170 175  
Asn Leu Leu Gln Ser Pro Leu Ser Ile Leu Leu Thr Leu Asp Val Glu  
180 185 190  
Leu Pro Pro Val Cys Cys Asp Ile Asp Ile  
195 200

<210> 384  
<211> 609  
<212> DNA  
<213> Homo sapiens

<400> 384  
atggctactc tgatctatgt tgataaggaa aacgaagaac caggcatcct tgtggctaca 60  
aaggatgggc tgaagctggg gtctggacct tcaatcaaag ccttagatgg gagatctcaa 120  
gtttcaatat catgttttgg caaaacattc gatgctccca catccttacc taaagctacc 180  
agaaaggctt tgggaactgt caacagagct acagaaaagt cagtaaagac caatggaccc 240  
ctcaaacaaa aacagccaag cttttctgcc aaaaagatga ctgagaagac tgttaaagca 300  
aaaaactctg ttcctgcctc agatgatggc tatccagaaa tagaaaaatt atttccttc 360  
aatcctctag gcttcgagag ttttgacctg cctgaagagc accagattgc acatctcccc 420  
ttgagtgaag tgcctctcat gatacttgat gaggagagag agcttgaaaa gctgtttcag 480  
ctgggcccc cttcaccttt gaagatgcc tctccaccat ggaaatccaa tctgttgacag 540  
tctcctttaa gcattctgtt gaccctggat gttgaattgc cacctgtttg ctgtgacata 600  
gatatttaa 609

<210> 385  
<211> 322  
<212> PRT  
<213> Homo sapiens

<400> 385

Met Glu Gly Ile Ser Asn Phe Lys Thr Pro Ser Lys Leu Ser Glu Lys  
 1 5 10 15  
 Lys Lys Ser Val Leu Cys Ser Thr Pro Thr Ile Asn Ile Pro Ala Ser  
 20 25 30  
 Pro Phe Met Gln Lys Leu Gly Phe Gly Thr Gly Val Asn Val Tyr Leu  
 35 40 45  
 Met Lys Arg Ser Pro Arg Gly Leu Ser His Ser Pro Trp Ala Val Lys  
 50 55 60  
 Lys Ile Asn Pro Ile Cys Asn Asp His Tyr Arg Ser Val Tyr Gln Lys  
 65 70 75 80  
 Arg Leu Met Asp Glu Ala Lys Ile Leu Lys Ser Leu His His Pro Asn  
 85 90 95  
 Ile Val Gly Tyr Arg Ala Phe Thr Glu Ala Asn Asp Gly Ser Leu Cys  
 100 105 110  
 Leu Ala Met Glu Tyr Gly Gly Glu Lys Ser Leu Asn Asp Leu Ile Glu  
 115 120 125  
 Glu Arg Tyr Lys Ala Ser Gln Asp Pro Phe Pro Ala Ala Ile Ile Leu  
 130 135 140  
 Lys Val Ala Leu Asn Met Ala Arg Gly Leu Lys Tyr Leu His Gln Glu  
 145 150 155 160  
 Lys Lys Leu Leu His Gly Asp Ile Lys Ser Ser Asn Val Val Ile Lys  
 165 170 175  
 Gly Asp Phe Glu Thr Ile Lys Ile Cys Asp Val Gly Val Ser Leu Pro  
 180 185 190  
 Leu Asp Glu Asn Met Thr Val Thr Asp Pro Glu Ala Cys Tyr Ile Gly  
 195 200 205  
 Thr Glu Pro Trp Lys Pro Lys Glu Ala Val Glu Glu Asn Gly Val Ile  
 210 215 220  
 Thr Asp Lys Ala Asp Ile Phe Ala Phe Gly Leu Thr Leu Trp Glu Met  
 225 230 235 240  
 Met Thr Leu Ser Ile Pro His Ile Asn Leu Ser Asn Asp Asp Asp  
 245 250 255  
 Glu Asp Lys Thr Phe Asp Glu Ser Asp Phe Asp Asp Glu Ala Tyr Tyr  
 260 265 270  
 Ala Ala Leu Gly Thr Arg Pro Pro Ile Asn Met Glu Glu Leu Asp Glu  
 275 280 285  
 Ser Tyr Gln Lys Val Ile Glu Leu Phe Ser Val Cys Thr Asn Glu Asp  
 290 295 300  
 Pro Lys Asp Arg Pro Ser Ala Ala His Ile Val Glu Ala Leu Glu Thr  
 305 310 315 320  
 Asp Val

&lt;210&gt; 386

&lt;211&gt; 1899

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 386

agcgcgcgac tttttgaaag ccaggagggt tcgaattgca acggcagctg ccgggcggtat 60  
 gtgttggtgc tagaggcagc tgcagggtct cgctgggggc cgctcgggac caattttgaa 120  
 gaggtacttg gccacgactt attttcacct ccgacctttc cttccaggcg gtgagactct 180  
 ggactgagag tggctttcac aatggaaggg atcagtaatt tcaagacacc aagcaaatta 240  
 tcagaaaaaa agaaatctgt attatgttca actccaacta taaatatccc ggcctctccg 300  
 tttatgcaga agcttggtct tgggtactggg gtaaattgtgt acctaataagaa aagatctcca 360  
 agaggtttgt ctcattctcc ttgggctgta aaaaagatta atcctatatg taatgatcat 420  
 tatcgaagtg tgtatcaaaa gagactaatg gatgaagcta agattttgaa aagccttcat 480  
 catccaaaca ttgttggtta tcgtgctttt actgaagcca atgatggcag tctgtgtctt 540  
 gctatggaat atggaggtga aaagtctcta aatgacttaa tagaagaacg atataaagcc 600  
 agccaagatc cttttccagc agccataatt ttaaaagttg ctttgaatat ggcaagaggg 660

```

ttaaagtatc tgcaccaaga aaagaaactg cttcatggag acataaagtc ttcaaagtgt 720
gtaatttaaag gcgatttttga aacaattaaa atctgtgatg taggagtctc tctaccactg 780
gatgaaaata tgactgtgac tgaccctgag gcttggttaca ttggcacaga gccatggaaa 840
cccaaagaag ctgtggagga gaatggtgtt attactgaca aggagacat atttgccctt 900
ggccttactt tgtgggaaat gatgacttta tgcattccac acattaatct ttcaaagtat 960
gatgatgatg aagataaaac ttttgatgaa agtgattttg atgatgaagc atactatgca 1020
gcgttgggaa ctaggccacc tattaatatg gaagaactgg atgaatcata ccagaaagta 1080
attgaactct tctctgtatg cactaatgaa gaccctaaag atcgtccttc tgctgcacac 1140
attgttgaag ctctggaaac agatgtctag tgatcatctc agctgaagtg tggcttgctg 1200
aaataactgt ttattccaaa atattttacat agttactatc agtagttatt agactctaaa 1260
attggcatat ttgaggacca tagtttcttg ttaacatatg gataactatt tctaataatga 1320
aatatgctta tattggctat aagcacttgg aattgtactg ggttttctgt aaagttttag 1380
aaactagcta cataagtact ttgatactgc tcatgctgac ttaaaacact agcagtaaaa 1440
cgctgtaaac tgtaacatta aattgaatga ccattacttt tattaatgat ctttctttaa 1500
tattctatat tttaatggat ctactgacat tagcactttg tacagtacaa aataaagtct 1560
acatttgttt aaaacactga accttttgct gatgtgttta tcaaatgata actggaagct 1620
gaggagaata tgcctcaaaa agagtagctc cttggatact tcagactctg gttacagatt 1680
gtcttgatct cttggatctc ctcagatctt tggtttttgc tttaatattat taaatgtatt 1740
ttccatactg agtttaaaat ttattaattt gtaccttaag catttcccag ctgtgtaaaa 1800
acaataaaac tcaaatagga tgataaagaa taaaggacac tttgggtacc agaaaaaaa 1860
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1899

```

<210> 387  
 <211> 202  
 <212> PRT  
 <213> Homo sapiens

```

<400> 387
Met Ala Thr Leu Ile Tyr Val Asp Lys Glu Asn Gly Glu Pro Gly Thr
  1           5           10           15
Arg Val Val Ala Lys Asp Gly Leu Lys Leu Gly Ser Gly Pro Ser Ile
          20          25          30
Lys Ala Leu Asp Gly Arg Ser Gln Val Ser Thr Pro Arg Phe Gly Lys
          35          40          45
Thr Phe Asp Ala Pro Pro Ala Leu Pro Lys Ala Thr Arg Lys Ala Leu
          50          55          60
Gly Thr Val Asn Arg Ala Thr Glu Lys Ser Val Lys Thr Lys Gly Pro
          65          70          75          80
Leu Lys Gln Lys Gln Pro Ser Phe Ser Ala Lys Lys Met Thr Glu Lys
          85          90          95
Thr Val Lys Ala Lys Ser Ser Val Pro Ala Ser Asp Asp Ala Tyr Pro
          100         105         110
Glu Ile Glu Lys Phe Phe Pro Phe Asn Pro Leu Asp Phe Glu Ser Phe
          115         120         125
Asp Leu Pro Glu Glu His Gln Ile Ala His Leu Pro Leu Ser Gly Val
          130         135         140
Pro Leu Met Ile Leu Asp Glu Glu Arg Glu Leu Glu Lys Leu Phe Gln
          145         150         155         160
Leu Gly Pro Pro Ser Pro Val Lys Met Pro Ser Pro Pro Trp Glu Ser
          165         170         175
Asn Leu Leu Gln Ser Pro Ser Ser Ile Leu Ser Thr Leu Asp Val Glu
          180         185         190
Leu Pro Pro Val Cys Cys Asp Ile Asp Ile
          195         200

```

<210> 388  
 <211> 728  
 <212> DNA  
 <213> Homo sapiens

&lt;400&gt; 388

```

gcggcctcag atgaatgcgg ctgttaagac ctgcaataat ccagaatggc tactctgac 60
tatgttgata aggaaaatgg agaaccaggc acccgtgtgg ttgctaagga tgggctgaag 120
ctggggctctg gaccttcaat caaagccita gatgggagat ctcaagtttc aacaccacgt 180
tttggaacaaa cgttcgatgc cccaccagcc ttacctaaag ctactagaaa ggctttggga 240
actgtcaaca gagctacaga aaagtctgta aagaccaagg gaccctcaa acaaaaacag 300
ccaagctttt ctgccaaaaa gatgactgag aagactgtta aagcaaaaag ctctgttcct 360
gcctcagatg atgcctatcc agaaatagaa aaattctttc ccttcaatcc tctagacttt 420
gagagttttg acctgcctga agagcaccag attgcgcacc tccccttgag tggagtgcct 480
ctcatgatcc ttgacgagga gagagagctt gaaaagctgt ttcagctggg ccccccttca 540
cctgtgaaga tgccctctcc accatgggaa tccaatctgt tgcagtctcc ttcaagcatt 600
ctgtcgaccc tggatgttga attgccacct gtttgctgtg acatagatat ttaaatttct 660
tagtgcttca gagtttgtgt gtatttgtat taataaagca ttcttcaaca gaaaaaaaaa 720
aaaaaaaaa

```

&lt;210&gt; 389

&lt;211&gt; 221

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 389

```

Met Ser Glu Thr Ala Pro Leu Ala Pro Thr Ile Pro Ala Pro Ala Glu
1      5      10      15
Lys Thr Pro Val Lys Lys Lys Ala Lys Lys Ala Gly Ala Thr Ala Gly
20     25     30
Lys Arg Lys Ala Ser Gly Pro Pro Val Ser Glu Leu Ile Thr Lys Ala
35     40     45
Val Ala Ala Ser Lys Glu Arg Ser Gly Val Ser Leu Ala Ala Leu Lys
50     55     60
Lys Ala Leu Ala Ala Ala Gly Tyr Asp Val Glu Lys Asn Asn Ser Arg
65     70     75     80
Ile Lys Leu Gly Leu Lys Ser Leu Val Ser Lys Gly Thr Leu Val Gln
85     90     95
Thr Lys Gly Thr Gly Ala Ser Gly Ser Phe Lys Leu Asn Lys Lys Ala
100    105    110
Ala Ser Gly Glu Gly Lys Pro Lys Ala Lys Lys Ala Gly Ala Ala Lys
115    120    125
Pro Arg Lys Pro Ala Gly Ala Ala Lys Lys Pro Lys Lys Val Ala Gly
130    135    140
Ala Ala Thr Pro Lys Lys Ser Ile Lys Lys Thr Pro Lys Lys Val Lys
145    150    155    160
Lys Pro Ala Thr Ala Ala Gly Thr Lys Lys Val Ala Lys Ser Ala Lys
165    170    175
Lys Val Lys Thr Pro Gln Pro Lys Lys Ala Ala Lys Ser Pro Ala Lys
180    185    190
Ala Lys Ala Pro Lys Pro Lys Ala Ala Lys Pro Lys Ser Gly Lys Pro
195    200    205
Lys Val Thr Lys Ala Lys Lys Ala Ala Pro Lys Lys Lys
210    215    220

```

&lt;210&gt; 390

&lt;211&gt; 777

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 390

```

catgctgttc tgacagtttg agattactta ttgtcttttc tgggaagaca aaaacatgtc 60
ggagactgct ccacttgctc ctaccattcc tgcaccgcga gaaaaaacac ctgtgaagaa 120
aaaggcgaag aaggcaggcg caactgctgg gaaacgcaaa gcatccggac cccagtagtc 180
tgagcttata accaaggcag tggcagcttc taaggagcgc agcggcggtt ctctggccgc 240

```

```

gcttaagaaa gcgcttgccg ctgctggcta cgatgtagaa aaaaacaaca gccgtatcaa 300
gcttggcctc aagagcttgg tgagcaaagg tactctgggt cagaccaaag gtaccgggtgc 360
ttctggctcc ttcaaactca acaagaaagc ggcttccggg gaaggcaaac ccaaggccaa 420
aaaggctggc gcagccaagc ctaggaagcc tgctggggca gccaagaagc ccaagaaggt 480
ggctggcgcc gctaccccgga agaaaagcat caaaaagact cctaagaagg taaagaagcc 540
agcaaccgct gctgggacca agaaagtggc caagagtgcg aaaaagggtga aaacacctca 600
gccaaaaaaa gctgccaaga gtccagctaa ggccaaagcc cctaagccca agggcgccaa 660
gcctaagtcg ggggaagccga aggttacaaa ggcaaagaag gcagctccga agaaaaagtg 720
aaactggcgg gacgttcccc ttgaaaatt ttaaacggct cttttcagag ccaccca 777

```

&lt;210&gt; 391

&lt;211&gt; 846

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 391

```

Met Ser Ser Ser His Phe Ala Ser Arg His Arg Lys Asp Ile Ser Thr
 1          5          10          15
Glu Met Ile Arg Thr Lys Ile Ala His Arg Lys Ser Leu Ser Gln Lys
      20          25          30
Glu Asn Arg His Lys Glu Tyr Glu Arg Asn Arg His Phe Gly Leu Lys
      35          40          45
Asp Val Asn Ile Pro Thr Leu Glu Gly Arg Ile Leu Val Glu Leu Asp
      50          55          60
Glu Thr Ser Gln Glu Leu Val Pro Glu Lys Thr Asn Val Lys Pro Arg
      65          70          75          80
Ala Met Lys Thr Ile Leu Gly Asp Gln Arg Lys Gln Met Leu Gln Lys
      85          90          95
Tyr Lys Glu Glu Lys Gln Leu Gln Lys Leu Lys Glu Gln Arg Glu Lys
      100          105          110
Ala Lys Arg Gly Ile Phe Lys Val Gly Arg Tyr Arg Pro Asp Met Pro
      115          120          125
Cys Phe Leu Leu Ser Asn Gln Asn Ala Val Lys Ala Glu Pro Lys Lys
      130          135          140
Ala Ile Pro Ser Ser Val Arg Ile Thr Arg Ser Lys Ala Lys Asp Gln
      145          150          155          160
Met Glu Gln Thr Lys Ile Asp Asn Glu Ser Asp Val Arg Ala Ile Arg
      165          170          175
Pro Gly Pro Arg Gln Thr Ser Glu Lys Lys Val Ser Asp Lys Glu Lys
      180          185          190
Lys Val Val Gln Pro Val Met Pro Thr Ser Leu Arg Met Thr Arg Ser
      195          200          205
Ala Thr Gln Ala Ala Lys Gln Val Pro Arg Thr Val Ser Ser Thr Thr
      210          215          220
Ala Arg Lys Pro Val Thr Arg Ala Ala Asn Glu Asn Glu Pro Glu Gly
      225          230          235          240
Lys Val Pro Ser Lys Gly Arg Pro Ala Lys Asn Val Glu Thr Lys Pro
      245          250          255
Asp Lys Gly Ile Ser Cys Lys Val Asp Ser Glu Glu Asn Thr Leu Asn
      260          265          270
Ser Gln Thr Asn Ala Thr Ser Gly Met Asn Pro Asp Gly Val Leu Ser
      275          280          285
Lys Met Glu Asn Leu Pro Glu Ile Asn Thr Ala Lys Ile Lys Gly Lys
      290          295          300
Asn Ser Phe Ala Pro Lys Asp Phe Met Phe Gln Pro Leu Asp Gly Leu
      305          310          315          320
Lys Thr Tyr Gln Val Thr Pro Met Thr Pro Arg Ser Ala Asn Ala Phe
      325          330          335
Leu Thr Pro Ser Tyr Thr Trp Thr Pro Leu Lys Thr Glu Val Asp Glu
      340          345          350
Ser Gln Ala Thr Lys Glu Ile Leu Ala Gln Lys Cys Lys Thr Tyr Ser

```

355	360	365
Thr Lys Thr Ile Gln Gln Asp Ser Asn Lys Leu Pro Cys Pro Leu Gly		
370	375	380
Pro Leu Thr Val Trp His Glu Glu His Val Leu Asn Lys Asn Glu Ala		
385	390	395
Thr Thr Lys Asn Leu Asn Gly Leu Pro Ile Lys Glu Val Pro Ser Leu		
405	410	415
Glu Arg Asn Glu Gly Arg Ile Ala Gln Pro His His Gly Val Pro Tyr		
420	425	430
Phe Arg Asn Ile Leu Gln Ser Glu Thr Glu Lys Leu Thr Ser His Cys		
435	440	445
Phe Glu Trp Asp Arg Lys Leu Glu Leu Asp Ile Pro Asp Asp Ala Lys		
450	455	460
Asp Leu Ile Arg Thr Ala Val Gly Gln Thr Arg Leu Leu Met Lys Glu		
465	470	475
Arg Phe Lys Gln Phe Glu Gly Leu Val Asp Asp Cys Glu Tyr Lys Arg		
485	490	495
Gly Ile Lys Glu Thr Thr Cys Thr Asp Leu Asp Gly Phe Trp Asp Met		
500	505	510
Val Ser Phe Gln Ile Glu Asp Val Ile His Lys Phe Asn Asn Leu Ile		
515	520	525
Lys Leu Glu Glu Ser Gly Trp Gln Val Asn Asn Asn Met Asn His Asn		
530	535	540
Met Asn Lys Asn Val Phe Arg Lys Lys Val Val Ser Gly Ile Ala Ser		
545	550	555
Lys Pro Lys Gln Asp Asp Ala Gly Arg Ile Ala Ala Arg Asn Arg Leu		
565	570	575
Ala Ala Ile Lys Asn Ala Met Arg Glu Arg Ile Arg Gln Glu Glu Cys		
580	585	590
Ala Glu Thr Ala Val Ser Val Ile Pro Lys Glu Val Asp Lys Ile Val		
595	600	605
Phe Asp Ala Gly Phe Phe Arg Val Glu Ser Pro Val Lys Leu Phe Ser		
610	615	620
Gly Leu Ser Val Ser Ser Glu Gly Pro Ser Gln Arg Leu Gly Thr Pro		
625	630	635
Lys Ser Val Asn Lys Ala Val Ser Gln Ser Arg Asn Glu Met Gly Ile		
645	650	655
Pro Gln Gln Thr Thr Ser Pro Glu Asn Ala Gly Pro Gln Asn Thr Lys		
660	665	670
Ser Glu His Val Lys Lys Thr Leu Phe Leu Ser Ile Pro Glu Ser Arg		
675	680	685
Ser Ser Ile Glu Asp Ala Gln Cys Pro Gly Leu Pro Asp Leu Ile Glu		
690	695	700
Glu Asn His Val Val Asn Lys Thr Asp Leu Lys Val Asp Cys Leu Ser		
705	710	715
Ser Glu Arg Met Ser Leu Pro Leu Leu Ala Gly Gly Val Ala Asp Asp		
725	730	735
Ile Asn Thr Asn Lys Lys Glu Gly Ile Ser Asp Val Val Glu Gly Met		
740	745	750
Glu Leu Asn Ser Ser Ile Thr Ser Gln Asp Val Leu Met Ser Ser Pro		
755	760	765
Glu Lys Asn Thr Ala Ser Gln Asn Ser Ile Leu Glu Glu Gly Glu Thr		
770	775	780
Lys Ile Ser Gln Ser Glu Leu Phe Asp Asn Lys Ser Leu Thr Thr Glu		
785	790	795
Cys His Leu Leu Asp Ser Pro Gly Leu Asn Cys Ser Asn Pro Phe Thr		
805	810	815
Gln Leu Glu Arg Arg His Gln Glu His Ala Arg His Ile Ser Phe Gly		
820	825	830
Gly Asn Leu Ile Thr Phe Ser Pro Leu Gln Pro Gly Glu Phe		
835	840	845

<210> 392  
<211> 2979  
<212> DNA  
<213> Homo sapiens

<400> 392

```
agcaaaccac tgcgaagcct cggtgagtg aaggggtgg atcttcccc gaagttttgg 60
ttaagcccc tccaatcagc ggctcggg gcgaagttt aatttcgtg aggcctcggg 120
tgtgaggggt cctgcttcgg agtcggcggt ggctcgtcc accgagtggt ctttactttt 180
tggttggttg aggtttcacg ctagaaggtg gctcaggatg tcttcatcac attttgccag 240
tcgacacagg aaggatataa gtactgaaat gattagaact aaaattgctc ataggaaatc 300
actgtctcag aaagaaaata gacataagga atacgaacga aatagacact ttggtttgaa 360
agatgtaaac attccaacct tggaaggtag aattcttggt gaattagatg agacatctca 420
agagcttggt ccagaaaaga ccaatgttaa gccaaaggca atgaaaacta ttctaggtga 480
tcaacgaaaa cagatgctcc aaaaatacaa agaagaaaag caacttcaaa aattgaaaga 540
gcagagagag aaagctaaac gaggaatatt taaagtgggt cgttatagac ctgatatgcc 600
ttgttttctt ttatcaaacc agaattgctgt gaaagctgag ccaaaaaagg ctattccatc 660
ttctgtacgg attacaaggt caaaggccaa agaccaaatg gagcagacta agattgataa 720
cgagagtgat gttcgagcaa tccgacctgg tccaagacaa acttctgaaa agaaagtgtc 780
agacaaagag aaaaaagttg tgcagcctgt aatgcccacg tcgttgagaa tgactcgatc 840
agctactcaa gcagcaaagc aggttcccag aacagtctca tctaccacag caagaaagcc 900
agtcacaaga gctgctaatt aaaacgaacc agaaggaaag gtgccaagta aaggaagacc 960
tgccaaaaat gtagaaacaa aacccgacaa gggattttct tgtaaagtcg atagtgaaga 1020
aaatactttg aattcacaaa ctaatgcaac aagtggaaat aatccagatg gagtcttatc 1080
aaaaattgaa aacttacctg agataaatac tgcaaaaata aaagggaaga attccttcgc 1140
acctaaggat tttatgtttc agccactgga tggctggaag acctatcaag taacacctat 1200
gactcccaga agtgccaatg cttttttgac acccagttac acctggactc ctttaaaaaa 1260
agaagttgat gagtctcaag caacaaaaga aattttggca caaaaatgta aaacttactc 1320
taccaagaca atacagcaag attcaaataa attgcatgtt cttttgggtc ctctaactgt 1380
ttggcatgaa gaacatgttt taaataaaaa tgaagctact actaaaaatt taaatggcct 1440
tccaataaaa gaagtcccat cacttgaaag aatgaaggt cgaattgctc agccccacca 1500
tggtgtgcca tatttcagaa atatcctcca gtcagaaact gagaaattaa cttcacattg 1560
cttcagtggt gacaggaaac ttgaattgga cattccagat gatgctaaag atcttattcg 1620
cacagcagtt ggtcaacaaa gactccttat gaaggaaagg tttaaacagt ttgaaggact 1680
ggttgatgat tgtgaatata aacgaggtat aaaggagact acctgtacag atctggatgg 1740
attttgggat atggttagtt ttcagataga agatgtaatc cacaaattca acaatctgat 1800
caaaacttgag gaatctgggt ggcaagtcaa taataatatg aatcataata tgaacaaaaa 1860
tgtcttttag aaaaaagttg tctcaggtat agcaagtaaa ccaaaacagg atgatgctgg 1920
aagaattgca gcgagaaatc gcctagctgc cataaaaaat gcaatgagag agagaattag 1980
gcaggaagaa tgtgctgaaa cagcagtttc tgtgatacca aagggaagtg ataaaaatag 2040
gttcgatgct ggatttttca gagttgaaag tcctgttaaa ttattctcag gactttctgt 2100
ctcttctgaa ggcccttctc aaagacttgg aacacctaag tctgtcaaca aagctgtatc 2160
tcagagtaga aatgagatgg gcattccaca acaaactaca tcaccagaaa atgccgggtc 2220
tcagaatacg aaaagtgaac atgtgaagaa gactttgttt ttgagtattc ctgaaagcag 2280
gagcagcata gaagatgctc agtgtcctgg attaccagat ttaattgaag aaaacctgt 2340
tgtaaaataa acagacttga aggtggattg tttatccagt gagagaatga gtttgccctc 2400
tcttgctggg ggagtagcag atgatattaa tactaacaac aaagaaggaa ttccagatgt 2460
tgtggaagga atggaactga attcttcaat tacatcacag gatgttttga tgagtagccc 2520
tgaaaaaaat acagcttcac aaaatagcat cttagaagaa ggggaaacta aaatttctca 2580
gtcagaacta tttgataata aaagtctcac tactgaatgc caccttcttg attcaccagg 2640
tctaaactgc agtaatccat ttactcagct ggagaggaga catcaagaac atgccagaca 2700
catttctttt ggtggaatc tgattacttt ttacctcta caaccaggag aattttgaat 2760
ttaaaaaata atccaaacat tttccttcat attatcaatg cttatatatt ccttagacta 2820
ttgaaatttt ggagaaaatg tatttgtgtt cacttctata gcatataatg ttttaattat 2880
ctgtgttcat caaagtgtat tttagatata ctcttctca aggggaagtg ggatattttg 2940
tacattttca acacagaata aaaaatgtac tgtgccttg 2979
```

<210> 393  
<211> 450

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 393

```

Met Asp Ala Arg Arg Val Pro Gln Lys Asp Leu Arg Val Lys Lys Asn
 1          5          10          15
Leu Lys Lys Phe Arg Tyr Val Lys Leu Ile Ser Met Glu Thr Ser Ser
 20          25          30
Ser Ser Asp Asp Ser Cys Asp Ser Phe Ala Ser Asp Asn Phe Ala Asn
 35          40          45
Thr Lys Pro Lys Phe Arg Ser Asp Ile Ser Glu Glu Leu Ala Ser Val
 50          55          60
Phe Tyr Glu Asp Ser Asp Asn Glu Ser Phe Cys Gly Phe Ser Glu Ser
 65          70          75          80
Glu Val Gln Asp Val Leu Asp His Cys Gly Phe Leu Gln Lys Pro Arg
 85          90          95
Pro Asp Val Thr Asn Glu Leu Ala Gly Ile Phe His Ala Asp Ser Asp
100          105          110
Asp Glu Ser Phe Cys Gly Phe Ser Glu Ser Glu Ile Gln Asp Gly Met
115          120          125
Arg Leu Gln Ser Val Arg Glu Gly Cys Arg Thr Arg Ser Gln Cys Arg
130          135          140
His Ser Gly Pro Leu Arg Val Ala Met Lys Phe Pro Ala Arg Ser Thr
145          150          155          160
Arg Gly Ala Thr Asn Lys Lys Ala Glu Ser Arg Gln Pro Ser Glu Asn
165          170          175
Ser Val Thr Asp Ser Asn Ser Asp Ser Glu Asp Glu Ser Gly Met Asn
180          185          190
Phe Leu Glu Lys Arg Ala Leu Asn Ile Lys Gln Asn Lys Ala Met Leu
195          200          205
Ala Lys Leu Met Ser Glu Leu Glu Ser Phe Pro Gly Ser Phe Arg Gly
210          215          220
Arg His Pro Leu Pro Gly Ser Asp Ser Gln Ser Arg Arg Pro Arg Arg
225          230          235          240
Arg Thr Phe Pro Gly Val Ala Ser Arg Arg Asn Pro Glu Arg Arg Ala
245          250          255
Arg Pro Leu Thr Arg Ser Arg Ser Arg Ile Leu Gly Ser Leu Asp Ala
260          265          270
Leu Pro Met Glu Glu Glu Glu Glu Glu Asp Lys Tyr Met Leu Val Arg
275          280          285
Lys Arg Lys Thr Val Asp Gly Tyr Met Asn Glu Asp Asp Leu Pro Arg
290          295          300
Ser Arg Arg Ser Arg Ser Ser Val Thr Leu Pro His Ile Ile Arg Pro
305          310          315          320
Val Glu Glu Ile Thr Glu Glu Glu Leu Glu Asn Val Cys Ser Asn Ser
325          330          335
Arg Glu Lys Ile Tyr Asn Arg Ser Leu Gly Ser Thr Cys His Gln Cys
340          345          350
Arg Gln Lys Thr Ile Asp Thr Lys Thr Asn Cys Arg Asn Pro Asp Cys
355          360          365
Trp Gly Val Arg Gly Gln Phe Cys Gly Pro Cys Leu Arg Asn Arg Tyr
370          375          380
Gly Glu Glu Val Arg Asp Ala Leu Leu Asp Pro Asn Trp His Cys Pro
385          390          395          400
Pro Cys Arg Gly Ile Cys Asn Cys Ser Phe Cys Arg Gln Arg Asp Gly
405          410          415
Arg Cys Ala Thr Gly Val Leu Val Tyr Leu Ala Lys Tyr His Gly Phe
420          425          430
Gly Asn Val His Ala Tyr Leu Lys Ser Leu Lys Gln Glu Phe Glu Met
435          440          445
Gln Ala

```



450

<210> 394  
<211> 2824  
<212> DNA  
<213> Homo sapiens

&lt;400&gt; 394

```
cccagacccc gccctccg gcccgggtcg ggcgcccag cctgccagcc gcgctgctgc 60
tgctcctcct gctgtgggac cgctgaccgc gcggtgctc cgctctcccc gctccaagcg 120
ccgatctggg caccgcccac cagcatggac gctcgccgcg tgccgcagaa agatctcaga 180
gtaaagaaga acttaaagaa attcagatat gtgaagtga tttccatgga aacctcgta 240
tcctctgatg acagtgtgta cagctttgct tctgataatt ttgcaaacac gaaacctaaa 300
ttcaggtcag atatcagtga agaactggca agtggttttt atgaggactc tgataatgaa 360
tctttctgcg gcttttcaga aagtgggtg caagatgtat tagaccattg tggattttta 420
cagaaaccaa ggccagatgt cactaacgaa ctggccggtg tttttcatgc cgactctgac 480
gatgaatcat tttgcggttt ctgagagagt gagatacaag atggaatgag gctgcagtca 540
gttcgggaag gctgtaggac ccgcagccag tgcaggcact ctggacctct cagggtggcg 600
atgaagtttc cagcgcgag taccagggga gcaaccaaca aaaaagcaga gtcccgcag 660
ccctcagaga attctgtgac tgattccaac tccgattcag aagatgaaa tggaatgaat 720
tttttgaga aaagggttt aaatataaac caaaacaaag caatgcttgc aaaactcatg 780
tctgaattag aaagcttccc tggctcggtc cgtggaagac atccccctcc aggctccgag 840
tcacaatcaa ggagaccgag aaggcgta tcccggtg ttgcttccag gagaaacct 900
gaacggagag ctgctcctct taccaggtca aggtcccga tcctcggtc ccttgacgct 960
ctacccatgg aggaggagga ggaagaggat aagtacatgt tggtagaaa gaggaagacc 1020
gtggatggct acatgaatga agatgacctg ccagaaagcc gtcgctccag atcatccgtg 1080
acccttccgc atataattcg ccagtgga gaaattacag aggaggagt ggagaacgct 1140
tgcagcaatt ctcgagagaa gatataaac ggttcactg gctctactg tcatcaatgc 1200
cgtcagaaga ctattgatac caaaacaaac tgcagaaacc cagactgctg gggcggtcga 1260
ggccagttct gtggcccctg ccttcgaaac cgttatggtg aagaggtcag ggatgctctg 1320
ctggatccga actggcattg ccgccttgt cgaggaatct gcaactgcag tttctgccg 1380
cagcgagatg gacggtgtgc gactgggtc cttgtgtatt tagccaaata tcatggcttt 1440
gggaatgtgc atgcctactt gaaaagcctg aaacaggaat ttgaaatgca agcataatat 1500
ctggaaaatt tgctgcctgc cttctacttc tcaaatctt cttgtaaaag tttccaattt 1560
tttactgaa acctgagtta aaaatcttga tgcagcct gtttcataag aaactccaat 1620
caagttaatc ttagcagaca tgtgtttctg gagcatcaca gaaggtatat tgctagtta 1680
actttgccc cctgcagttt cttctctgct cccaaccccc atctcatagc atccccctct 1740
atttccaatg ctctctcca accgcttagt ttctgaattt cttttaaatt acagttttat 1800
gaaagcatat tttatttact tgggtgtgaa atagccctca taaaacctaa gcacttgga 1860
acacaataat agtattaact aactagatct attgaatttc agagaagagc cttctaactt 1920
gtttacacaa aaacgagtat gatttagcat tcatactagt tgaaattttt aatagaatca 1980
aggcacaaaa gtcttaaaac catgtggaaa aattaggtaa ttattgcaga ttgatgtctc 2040
tcaatcccat gtattgcgtt tatgttacaa gttgtgttca cagttgagac ttaatttctc 2100
ctaatttctt ctgcccgaag ggtaagtggg gcgtccagct tacacaatca taattcaaag 2160
gttggtgggc aatgtaatac ttaattaaaa taatgatgga agagctatct ggagattatg 2220
agtaagctga tttgaatttt cagtataaaa ctttagtata attgtagttt gcaaagttta 2280
tttcagttca catgtaagggt attgcaataa aattcttgga caattttgta tggaaacttg 2340
atattaaaaa ctagtctgtg gttctttgca gtttcttgta aatttataaa ccaggcacaa 2400
ggttcaagtt tagattttta gcacttttat aacaatgata agtgcccttt tggagatgta 2460
acttttagca gtttgtaaac ctgacatctc tgccagtcta gtttctgggc aggtttcctg 2520
tgctcagttt cccctcctc tttgcattaa tcaaggtatt tggtagaggt ggaatctaag 2580
tgtttgatg tccaatttac ttgcatatgt aaaccattgc tgtgccattc aatgtttgat 2640
gcataattgg accttgaatc gataagtgt aatacagctt ttgatctgta atgcttttat 2700
acaaaagttt attttaataa taaaatgttt gttctaactt gtctgctttt ttaaaaataa 2760
tcttactgta ctttaattcta attttttctt catattttaa taaaaggcca tttccacctt 2820
ttct
```

<210> 395  
<211> 142  
<212> PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 395

Met Gly Ala Pro Thr Leu Pro Pro Ala Trp Gln Pro Phe Leu Lys Asp  
 1 5 10 15  
 His Arg Ile Ser Thr Phe Lys Asn Trp Pro Phe Leu Glu Gly Cys Ala  
 20 25 30  
 Cys Thr Pro Glu Arg Met Ala Glu Ala Gly Phe Ile His Cys Pro Thr  
 35 40 45  
 Glu Asn Glu Pro Asp Leu Ala Gln Cys Phe Phe Cys Phe Lys Glu Leu  
 50 55 60  
 Glu Gly Trp Glu Pro Asp Asp Asp Pro Ile Glu Glu His Lys Lys His  
 65 70 75 80  
 Ser Ser Gly Cys Ala Phe Leu Ser Val Lys Lys Gln Phe Glu Glu Leu  
 85 90 95  
 Thr Leu Gly Glu Phe Leu Lys Leu Asp Arg Glu Arg Ala Lys Asn Lys  
 100 105 110  
 Ile Ala Lys Glu Thr Asn Asn Lys Lys Lys Glu Phe Glu Glu Thr Ala  
 115 120 125  
 Lys Lys Val Arg Arg Ala Ile Glu Gln Leu Ala Ala Met Asp  
 130 135 140

&lt;210&gt; 396

&lt;211&gt; 1619

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 396

ccgccagatt tgaatcgagg gacccgttgg cagaggtggc ggccggcgga tgggtgcccc 60  
 gacgttgccc cctgcctggc agccctttct caaggaccac cgcattctcta cattcaagaa 120  
 ctggcccttc ttggagggtc gcgcctgcac cccggagcgg atggccgagg ctggcttcat 180  
 ccaactgcccc actgagaacg agccagactt ggcccagtggt ttcttctgct tcaaggagct 240  
 ggaaggctgg gagccagatg acgaccccat agaggaacat aaaaagcatt cgtccggttg 300  
 cgcttttctt tctgtcaaga agcagtttga agaattaacc cttggtgaat ttttgaaact 360  
 ggacagagaa agagccaaga acaaaattgc aaaggaaacc aacaataaga agaaagaatt 420  
 tgaggaaact gcgaagaaag tgcgcctgct catcgagcag ctggctgcca tggattgagg 480  
 cctctggccg gagctgcctg gtcccagagt ggtgcacca cttccagggt ttattccctg 540  
 gtgccaccag ctttctctgt ggccccttag caatgtctta ggaaaggaga tcaacatttt 600  
 caaattagat gtttcaactg tgctcctgtt ttgtcttgaa agtggcacca gaggtgcttc 660  
 tgccctgtgca gcgggtgctg ctggtaacag tggtgtcttc tctctctctc tctctttttt 720  
 gggggctcat ttttgctgtt ttgattcccg ggcttaccag gtgagaagtg agggaggaag 780  
 aaggcagtggt cctttttgct agagctgaca gctttgttcg cgtgggcaga gccttcacca 840  
 gtgaatgtgt ctggacctca tgttgttgag gctgtcacag tcttgagtgt ggacttgga 900  
 ggtgcctgtt gaatctgagc tgcaggttcc ttatctgtca cacctgtgcc tctcagagg 960  
 acagtttttt tgttgttggt tttttttgtt tttttttttt ggtagatgca tgacttggtg 1020  
 gtgatgagag aatggagaca ggtccctgg ctctctact gtttaacaac atggctttct 1080  
 tattttgttt gaattgttaa ttcacagaat agcacaact acaattaaaa ctaagcaca 1140  
 agccattcta agtcattggg gaaacggggg gaacttcagg tggatgagga gacagaatag 1200  
 agtgatagga agcgtctggc agatactcct ttgtccactg ctgtgtgatt agacaggccc 1260  
 agtgagccgc ggggcacatg ctggcogctc ctccctcaga aaaaggcagt ggcctaaatc 1320  
 ctttttaaat gacttggctc gatgctgtgg gggactggct gggctgctgc aggcgtgtg 1380  
 tctgtcagcc caaccttcac atctgtcacg ttctccacac gggggagaga cgcagtcgc 1440  
 ccaggtcccc gctttctttg gaggcagcag ctcccgagg gctgaagtct ggcgtaagat 1500  
 gatggatttg attcgccctc ctccctgtca tagagctgca ggggtgattg ttacagcttc 1560  
 gctggaaacc tctggagggtc atctcggtgt ttctgagaa ataaaaagcc tgtcatttc 1619

&lt;210&gt; 397

&lt;211&gt; 1401

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

<400> 397  
Met Ala Pro Gln Met Tyr Glu Phe His Leu Pro Leu Ser Pro Glu Glu  
1 5 10 15  
Leu Leu Lys Ser Gly Gly Val Asn Gln Tyr Val Val Gln Glu Val Leu  
20 25 30  
Ser Ile Lys His Leu Pro Pro Gln Leu Arg Ala Phe Gln Ala Ala Phe  
35 40 45  
Arg Ala Gln Gly Pro Leu Ala Met Leu Gln His Phe Asp Thr Ile Tyr  
50 55 60  
Ser Ile Leu His His Phe Arg Ser Ile Asp Pro Gly Leu Lys Glu Asp  
65 70 75 80  
Thr Leu Glu Phe Leu Ile Lys Val Val Ser Arg His Ser Gln Glu Leu  
85 90 95  
Pro Ala Ile Leu Asp Asp Thr Thr Leu Ser Gly Ser Asp Arg Asn Ala  
100 105 110  
His Leu Asn Ala Leu Lys Met Asn Cys Tyr Ala Leu Ile Arg Leu Leu  
115 120 125  
Glu Ser Phe Glu Thr Met Ala Ser Gln Thr Asn Leu Val Asp Leu Asp  
130 135 140  
Leu Gly Gly Lys Gly Lys Lys Ala Arg Thr Lys Ala Ala His Gly Phe  
145 150 155 160  
Asp Trp Glu Glu Glu Arg Gln Pro Ile Leu Gln Leu Leu Thr Gln Leu  
165 170 175  
Leu Gln Leu Asp Ile Arg His Leu Trp Asn His Ser Ile Ile Glu Glu  
180 185 190  
Glu Phe Val Ser Leu Val Thr Gly Cys Cys Tyr Arg Leu Leu Glu Asn  
195 200 205  
Pro Thr Ile Asn His Gln Lys Asn Arg Pro Thr Arg Glu Ala Ile Thr  
210 215 220  
His Leu Leu Gly Val Ala Leu Thr Arg Tyr Asn His Met Leu Ser Ala  
225 230 235 240  
Thr Val Lys Ile Ile Gln Met Leu Gln His Phe Glu His Leu Ala Pro  
245 250 255  
Val Leu Val Ala Ala Val Ser Leu Trp Ala Thr Asp Tyr Gly Met Lys  
260 265 270  
Ser Ile Val Gly Glu Ile Val Arg Glu Ile Gly Gln Lys Cys Pro Gln  
275 280 285  
Glu Leu Ser Arg Asp Pro Ser Gly Thr Lys Gly Phe Ala Ala Phe Leu  
290 295 300  
Thr Glu Leu Ala Glu Arg Val Pro Ala Ile Leu Met Ser Ser Met Cys  
305 310 315 320  
Ile Leu Leu Asp His Leu Asp Gly Glu Asn Tyr Met Met Arg Asn Ala  
325 330 335  
Val Leu Ala Ala Met Ala Glu Met Val Leu Gln Val Leu Ser Gly Asp  
340 345 350  
Gln Leu Glu Ala Ala Ala Arg Asp Thr Arg Asp Gln Phe Leu Asp Thr  
355 360 365  
Leu Gln Ala His Gly His Asp Val Asn Ser Phe Val Arg Ser Arg Val  
370 375 380  
Leu Gln Leu Phe Thr Arg Ile Val Gln Gln Lys Ala Leu Pro Leu Thr  
385 390 395 400  
Arg Phe Gln Ala Val Val Ala Leu Ala Val Gly Arg Leu Ala Asp Lys  
405 410 415  
Ser Val Leu Val Cys Lys Asn Ala Ile Gln Leu Leu Ala Ser Phe Leu  
420 425 430  
Ala Asn Asn Pro Phe Ser Cys Lys Leu Ser Asp Ala Asp Leu Ala Gly  
435 440 445  
Pro Leu Gln Lys Glu Thr Gln Lys Leu Gln Glu Met Arg Ala Gln Arg  
450 455 460  
Arg Thr Ala Ala Ala Ser Ala Val Leu Asp Pro Glu Glu Glu Trp Glu

465		470		475		480
Ala Met Leu Pro Glu	Leu Lys Ser Thr	Leu Gln Gln Leu Leu Gln Leu				
	485	490	495			
Pro Gln Gly Glu Glu Ile Pro	Glu Gln Ile Ala Asn Thr Glu Thr					
	500	505	510			
Thr Glu Asp Val Lys Gly Arg Ile Tyr	Gln Leu Leu Ala Lys Ala Ser					
	515	520	525			
Tyr Lys Lys Ala Ile Ile Leu Thr Arg	Glu Ala Thr Gly His Phe Gln					
	530	535	540			
Glu Ser Glu Pro Phe Ser His Ile Asp	Pro Glu Ser Glu Glu Thr					
545	550	555	560			
Arg Leu Leu Asn Ile Leu Gly Leu Ile	Phe Lys Gly Pro Ala Ala Ser					
	565	570	575			
Thr Gln Glu Lys Asn Pro Arg Glu Ser	Thr Gly Asn Met Val Thr Gly					
	580	585	590			
Gln Thr Val Cys Lys Asn Lys Pro Asn	Met Ser Asp Pro Glu Glu Ser					
	595	600	605			
Arg Gly Asn Asp Glu Leu Val Lys Gln	Glu Met Leu Val Gln Tyr Leu					
610	615	620				
Gln Asp Ala Tyr Ser Phe Ser Arg Lys	Ile Thr Glu Ala Ile Gly Ile					
625	630	635	640			
Ile Ser Lys Met Met Tyr Glu Asn Thr	Thr Thr Val Val Gln Glu Val					
	645	650	655			
Ile Glu Phe Phe Val Met Val Phe Gln	Phe Gly Val Pro Gln Ala Leu					
	660	665	670			
Phe Gly Val Arg Arg Met Leu Pro Leu	Ile Trp Ser Lys Glu Pro Gly					
	675	680	685			
Val Arg Glu Ala Val Leu Asn Ala Tyr	Arg Gln Leu Tyr Leu Asn Pro					
690	695	700				
Lys Gly Asp Ser Ala Arg Ala Lys Ala	Gln Ala Leu Ile Gln Asn Leu					
705	710	715	720			
Ser Leu Leu Leu Val Asp Ala Ser Val	Gly Thr Ile Gln Cys Leu Glu					
	725	730	735			
Glu Ile Leu Cys Glu Phe Val Gln Lys	Asp Glu Leu Lys Pro Ala Val					
	740	745	750			
Thr Gln Leu Leu Trp Glu Arg Ala Thr	Glu Lys Val Ala Cys Cys Pro					
	755	760	765			
Leu Glu Arg Cys Ser Ser Val Met Leu	Leu Gly Met Met Ala Arg Gly					
770	775	780				
Lys Pro Glu Ile Val Gly Ser Asn Leu	Asp Thr Leu Val Ser Ile Gly					
785	790	795	800			
Leu Asp Glu Lys Phe Pro Gln Asp Tyr	Arg Leu Ala Gln Gln Val Cys					
	805	810	815			
His Ala Ile Ala Asn Ile Ser Asp Arg	Arg Lys Pro Ser Leu Gly Lys					
	820	825	830			
Arg His Pro Pro Phe Arg Leu Pro Gln	Glu His Arg Leu Phe Glu Arg					
	835	840	845			
Leu Arg Glu Thr Val Thr Lys Gly Phe	Val His Pro Asp Pro Leu Trp					
850	855	860				
Ile Pro Phe Lys Glu Val Ala Val Thr	Leu Ile Tyr Gln Leu Ala Glu					
865	870	875	880			
Gly Pro Glu Val Ile Cys Ala Gln Ile	Leu Gln Gly Cys Ala Lys Gln					
	885	890	895			
Ala Leu Glu Lys Leu Glu Glu Lys Arg	Thr Ser Gln Glu Asp Pro Lys					
	900	905	910			
Glu Ser Pro Ala Met Leu Pro Thr Phe	Leu Leu Met Asn Leu Leu Ser					
	915	920	925			
Leu Ala Gly Asp Val Ala Leu Gln Gln	Leu Val His Leu Glu Gln Ala					
930	935	940				
Val Ser Gly Glu Leu Cys Arg Arg Arg	Val Leu Arg Glu Glu Gln Glu					
945	950	955	960			

His Lys Thr Lys Asp Pro Lys Glu Lys Asn Thr Ser Ser Glu Thr Thr  
 965 970 975  
 Met Glu Glu Glu Leu Gly Leu Val Gly Ala Thr Ala Asp Asp Thr Glu  
 980 985 990  
 Ala Glu Leu Ile Arg Gly Ile Cys Glu Met Glu Leu Leu Asp Gly Lys  
 995 1000 1005  
 Gln Thr Leu Ala Ala Phe Val Pro Leu Leu Leu Lys Val Cys Asn Asn  
 1010 1015 1020  
 Pro Gly Leu Tyr Ser Asn Pro Asp Leu Ser Ala Ala Ala Ser Leu Ala  
 1025 1030 1035 1040  
 Leu Gly Lys Phe Cys Met Ile Ser Ala Thr Phe Cys Asp Ser Gln Leu  
 1045 1050 1055  
 Arg Leu Leu Phe Thr Ile Leu Glu Lys Ser Pro Leu Pro Ile Val Arg  
 1060 1065 1070  
 Ser Asn Leu Met Val Ala Thr Gly Asp Leu Ala Ile Arg Phe Pro Asn  
 1075 1080 1085  
 Leu Val Asp Pro Trp Thr Pro His Leu Tyr Ala Arg Leu Arg Asp Pro  
 1090 1095 1100  
 Ala Gln Gln Val Arg Lys Thr Ala Gly Leu Val Met Thr His Leu Ile  
 1105 1110 1115 1120  
 Leu Lys Asp Met Val Lys Val Lys Gly Gln Val Ser Glu Met Ala Val  
 1125 1130 1135  
 Leu Leu Ile Asp Pro Glu Pro Gln Ile Ala Ala Leu Ala Lys Asn Phe  
 1140 1145 1150  
 Phe Asn Glu Leu Ser His Lys Gly Asn Ala Ile Tyr Asn Leu Leu Pro  
 1155 1160 1165  
 Asp Ile Ile Ser Arg Leu Ser Asp Pro Glu Leu Gly Val Glu Glu Glu  
 1170 1175 1180  
 Pro Phe His Thr Ile Met Lys Gln Leu Leu Ser Tyr Ile Thr Lys Asp  
 1185 1190 1195 1200  
 Lys Gln Thr Glu Ser Leu Val Glu Lys Leu Cys Gln Arg Phe Arg Thr  
 1205 1210 1215  
 Ser Leu Thr Glu Arg Gln Gln Arg Asp Leu Ala Tyr Cys Val Ser Gln  
 1220 1225 1230  
 Leu Pro Leu Thr Glu Arg Gly Leu Arg Lys Met Leu Asp Asn Phe Asp  
 1235 1240 1245  
 Cys Phe Gly Asp Lys Leu Ser Asp Glu Ser Ile Phe Ser Ala Phe Leu  
 1250 1255 1260  
 Ser Val Val Gly Lys Leu Arg Arg Gly Ala Lys Pro Glu Gly Lys Ala  
 1265 1270 1275 1280  
 Ile Ile Asp Glu Phe Glu Gln Lys Leu Arg Ala Cys His Thr Arg Gly  
 1285 1290 1295  
 Leu Asp Gly Ile Lys Glu Leu Glu Ile Gly Gln Ala Gly Ser Gln Arg  
 1300 1305 1310  
 Ala Pro Ser Ala Lys Lys Pro Ser Thr Gly Ser Arg Tyr Gln Pro Leu  
 1315 1320 1325  
 Ala Ser Thr Ala Ser Asp Asn Asp Phe Val Thr Pro Glu Pro Arg Arg  
 1330 1335 1340  
 Thr Thr Arg Arg His Pro Asn Thr Gln Gln Arg Ala Ser Lys Lys Lys  
 1345 1350 1355 1360  
 Pro Lys Val Val Phe Ser Ser Asp Glu Ser Ser Glu Glu Asp Leu Ser  
 1365 1370 1375  
 Ala Glu Met Thr Glu Asp Glu Thr Pro Lys Lys Thr Thr Pro Ile Leu  
 1380 1385 1390  
 Arg Ala Ser Ala Arg Arg His Arg Ser  
 1395 1400

<210> 398  
 <211> 5547  
 <212> DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 398

```

aagtttttga ggcgggaca agctgaggtc cgcgactcgt cgctaagatt ccccaaaact 60
gagattttcaa gaaaattcac cctttccgct ttctttggcg cctttcaacc gtgaaggaaa 120
tgaaggttga gaacctggaa cccgcttcca aagaccaagc cttttctctg tgcctcacia 180
cagtcttgcg ctgtagctgt ttttattacc cctattttac tactggggag gagatttgag 240
gtcctgaaag tgaagtgact tgctggacgt cacttagcag aggagatggc gaaacctgga 300
ttagaggcga tgccttctaa ctcccagcgt ggacttgccct cttttcctgg ggggtgactga 360
atgccagacc agggacgcga cgtctctggc cagcagaaat acggcctcct ccccgccgac 420
tgggcaaagg ttgaaacttg ggccaaggag ggattcgag gcgggcccgg ggtgggagcg 480
ggggccggag ccggagcctt ggcccgcgcc cgggcggcgc tgtgattggc cggccgctcc 540
ggcggggcgtc gcgattggcc gtccctgcagc cgttgagatt tgaactcggc atttgtggct 600
ttgcccgccg gtgcccagac tcagaggcgg ccctttgcct ctgcctgccg gggattggcc 660
ggattctccg ccgacttgaa aactgccttg ctaattgggt cgtgttggtc acgctgtgtt 720
tttcttttct atttcagcct gactgcggga atcagagccg cgggtgagat cccagccct 780
gtgagcctgt agagtagaa tggctcccc aatgtatgag ttccatctgc cattatcccc 840
agaggagtgt ttgaaaagtg gaggggtgaa tcagtatgtt gtgcaagagg tactgtccat 900
caaacatctt ccaccacagc ttagagcttt tcaggctgcc ttctgagctc aggggccccct 960
ggctatgctg cagcactttg atactatcta cagcattttg catcactttc gaagtataga 1020
tcctggcctc aaagaagata ctctggaatt cctgataaaa gtggtatccc gccactcccc 1080
ggagcttcca gctatcctgg atgatacaac tttgagtggg tcagatagaa acgcccctct 1140
aaatgcctc aaaaatgaact gttatgctct gatacgtctc ctggaatcct ttgagaccat 1200
ggccagccag acaaaccttg tggacctgga ccttggtggg aagggtgaga aagctcggac 1260
caaggcagcc catggccttg actgggaaga agagaggcaa ccaattcttc agcttttaac 1320
acagctactt cagttggaca tccgtcacct gtggaaccac tcaataattg aagaagaatt 1380
tgtcagtttg gttactggct gttgctaccg cttcttgagg aatcccacca ttaatcacca 1440
gaagaaccgc cccactcggg aagccataac acacctgctt ggtgtagcct tgaccggtta 1500
taaccatatg ctcatgtcta cagtgaagat catccagatg ctgcagcact ttgaacacct 1560
ggcacctgta ctggttgagc ccgtgagctc atgggcaact gactatggaa tgaagagcat 1620
agtgggagag attgtaagag agattggaca aaagtgtccc caagagctga gtcgagacc 1680
ttcagggaca aagggctttg cagcattcct gcagaaacta gcagaacgtg tcccagctat 1740
cctgatgtcc agcatgtgca ttttgctaga tcacctggat ggagaaaatt acatgatgag 1800
taatgctgtg ctggcagcca tggcggagat ggtgctgcag gttctcagtg gcgatcaact 1860
ggaagcagca gcccagagca ccagagacca gttcttggtt actttacaag cccatggcca 1920
tgatgtcaac tcctttgtgc ggagccgtgt tttgcagctc ttcaccgcaa ttgtccagca 1980
gaaggctctc cccctgacac gtttccaggc agtgggtggc ttagctgtgg gacgtctggc 2040
agacaagtca gtgctaagt gtaaaaatgc catccagctg ctggccagtt ttctagccaa 2100
taatcctttc tcctgcaagc ttagtgatgc tgacctgccc ggaccactgc agaaggagac 2160
ccagaaatta caagagatga gggcccagag gcgaactgca gcagcttctg cagtgtctgga 2220
cccagaggag gagtgggaag ccatgctgcc agagttgaag tctaccctgc agcagcttct 2280
acagcttccc cagggagagg aggagattcc tgagcaaat gccaatacag agacaactga 2340
agatgtgaaa ggacgcact atcaactgct tgccaaagct agttacaaa aggccatcat 2400
tctcactcga gaagccacag gccacttcca ggagtccgaa ccttcagtc atatagacc 2460
agaggagtca gaggagacca ggtccttgaa tatattagga cttatcttca aaggccagc 2520
agcttccaca caagaaaaga atccccggga gtctacagga aacatggtca caggacagac 2580
tgtctgtaaa aataaaccga atatgtcgga tcctgaggaa tccaggggaa atgatgaact 2640
agtgaagcag gagatgctgg tacagtatct gcaggatgcc tacagcttct cccggaagat 2700
tacagaggcc attggcatca tcagcaagat gatgtatgaa aacacaacta cagtgggtgca 2760
ggaggtgatt gaattctttg tgatggtctt ccaatttggt gtaccccagg ccctgtttgg 2820
ggtgcgcctg atctgcctc tcactgtggt taaggagcct ggtgtccggg aagccgtgct 2880
taatgcctac cgccaactct acctcaacct caaaggggac tctgccagag ccaaggccca 2940
ggctttgatt cagaatctct ctctgctgct agtggatgcc tcggttgagg ccattcagtg 3000
tcttgaggaa attctctgtg agtttggtga gaaggatgag ttgaaaccag cagtgaacca 3060
gctgctgtgg gagcgggcca ccgagaaggt cgcctgctgt cctctggagc gctgttcctc 3120
tgtcatgctt cttggcatga tggcacgagg aaagccagaa attgtgggaa gcaatttaga 3180
cacactggtg agcatagggc tggatgagaa gtttccacag gactacaggc tggcccagca 3240
ggtgtgccat gccactcgca acatctcgga caggagaaa ccttctctgg gcaaactgca 3300
cccccccttc cggctgcctc aggaacacag gttgtttgag cgactgcggg agacagtcac 3360
aaaaggcttt gtccaccag acccactctg gatccattc aaagaggtgg cagtgaacct 3420
catttaccaa ctggcagagg gcccgaagt gatctgtgcc cagatattgc agggctgtgc 3480

```

```

aaaacaggcc ctggagaagc tagaagagaa gagaaccagt caggaggacc cgaaggagtc 3540
ccccgcaatg ctccccactt tcctgttgat gaacctgctg tccctggctg gggatgtggc 3600
tctgcagcag ctggtccact tggagcaggc agtgagtgga gagctctgcc ggcgccgagt 3660
tctccgggaa gaacaggagc acaagaccaa agatcccaag gagaagaata cgagctctga 3720
gaccaccatg gaggaggagc tggggctggt tggggcaaca gcagatgaca cagaggcaga 3780
actaatccgt ggcattctgcg agatggaact gttggatggc aaacagacac tggctgcctt 3840
tggtccactc ttgcttaaag tctgtaacaa cccaggcctc tatagcaacc cagacctctc 3900
tgcagctgct tcacttgccc ttggcaagtt ctgcatgata agtgccactt tctgcgactc 3960
ccagcttcgt cttctgttca ccattcctgga aaagtctcca cttcccatg tccgggtctaa 4020
cctcatgggt gccactgggg atctggccat ccgctttccc aatctggtgg acccctggac 4080
tcctcatctg tatgctcgcc tccgggaccc tgctcagcaa gtgcggaaaa cagcggggct 4140
ggtgatgacc cacctgatcc tcaaggacat ggtgaagggt aaggggcagg tcagcgagat 4200
ggcgggtctg ctcatcgacc ccgagcctca gattgctgcc ctggccaaga acttcttcaa 4260
tgagctctcc cacaagggca acgcaatcta taatctcctt ccagatatca tcagccgcct 4320
gtcagacccc gagctggggg tggaggaaga gcctttccac accatcatga aacagctcct 4380
ctcctacatc accaaggaca agcagacaga gaggctggtg gaaaagctgt gtcagcggtt 4440
ccgcacatcc ctaactgagc ggcagcagcg agacctggcc tactgtgtgt cacagctgcc 4500
cctcacagag cgaggcctcc gtaagatgct tgacaatttt gactgttttg gagacaaact 4560
gtcagatgag tccatcttca gtgctttttt gtcagttgta ggcaagctgc gacgtggggc 4620
caagcctgag ggcaaggcta taatagatga atttgagcag aagcttcggg cctgtcatac 4680
cagaggtttg gatggaatca aggagcttga gattggccaa gcaggtagcc agagagcgcc 4740
atcagccaag aaaccatcca ctggttctag gtaccagcct ctggcttcta cagcctcaga 4800
caatgacttt gtcacaccag agccccgccc tactaccctg cggcattcaa acaccagca 4860
gagagcttcc aaaaagaaac ccaaagttgt cttctcaagt gatgagtcca gtgaggaaga 4920
tctttcagca gagatgacag aagacgagac acccaagaaa acaactccca ttctcagagc 4980
atcggctcgc aggcacagat cctaggaagt ctgttccctg cctccctgtg cagggtatcc 5040
tgtagggtga cctggaattc gaattctgtt tcccttgtaa aatatttgtc tgtctctttt 5100
ttttaaaaaa aaaaaggcc gggcactgtg gctcacgcct gtaatcccag cactttgcga 5160
taccaaggcg ggtggataac ctgaggtagg gaggctcgaga ccagcctgac caacatggag 5220
aaaccccatc tctactaaaa ataaaaaatt agccgggcgt attggcgtgc gcctgtaatc 5280
ccagctactc aagaggctga ggcaggagaa tcgcctgaac ccagaggcgg aggttgtagt 5340
gagccgaaat cacaccattg cactccagct tgggcaacaa tagcgaacct ccatctcaaa 5400
ttaaaaaaaa aatgcctaca cgctctttaa aatgcaaggc tttctcttaa attagcctaa 5460
ctgaactgcg ttgagctgct tcaacttttg aatatatgtt tgccaatctc cttgttttct 5520
aatgaataaa tgtttttata tacttttt 5547

```

<210> 399  
 <211> 213  
 <212> PRT  
 <213> Homo sapiens

```

<400> 399
Met Val Gln Ser Cys Ser Ala Tyr Gly Cys Lys Asn Arg Tyr Asp Lys
 1           5           10           15
Asp Lys Pro Val Ser Phe His Lys Phe Pro Leu Thr Arg Pro Ser Leu
      20           25           30
Cys Lys Glu Trp Glu Ala Ala Val Arg Arg Lys Asn Phe Lys Pro Thr
      35           40           45
Lys Tyr Ser Ser Ile Cys Ser Glu His Phe Thr Pro Asp Cys Phe Lys
      50           55           60
Arg Glu Cys Asn Asn Lys Leu Leu Lys Glu Asn Ala Val Pro Thr Ile
      65           70           75           80
Phe Leu Cys Thr Glu Pro His Asp Lys Lys Glu Asp Leu Leu Glu Pro
      85           90           95
Gln Glu Gln Leu Pro Pro Pro Pro Leu Pro Pro Pro Val Ser Gln Val
      100          105          110
Asp Ala Ala Ile Gly Leu Leu Met Pro Pro Leu Gln Thr Pro Val Asn
      115          120          125
Leu Ser Val Phe Cys Asp His Asn Tyr Thr Val Glu Asp Thr Met His
      130          135          140
Gln Arg Lys Arg Ile His Gln Leu Glu Gln Gln Val Glu Lys Leu Arg

```

<400> 401



Met	Gly	Lys	Lys	Arg	Thr	Lys	Gly	Lys	Thr	Val	Pro	Ile	Asp	Asp	Ser	1	5	10	15
Ser	Glu	Thr	Leu	Glu	Pro	Val	Cys	Arg	His	Ile	Arg	Lys	Gly	Leu	Glu	20	25	30	
Gln	Gly	Asn	Leu	Lys	Lys	Ala	Leu	Val	Asn	Val	Glu	Trp	Asn	Ile	Cys	35	40	45	
Gln	Asp	Cys	Lys	Thr	Asp	Asn	Lys	Val	Lys	Asp	Lys	Ala	Glu	Glu	Glu	50	55	60	
Thr	Glu	Glu	Lys	Pro	Ser	Val	Trp	Leu	Cys	Leu	Lys	Cys	Gly	His	Gln	65	70	75	80
Gly	Cys	Gly	Arg	Asn	Ser	Gln	Glu	Gln	His	Ala	Leu	Lys	His	Tyr	Leu	85	90	95	
Thr	Pro	Arg	Ser	Glu	Pro	His	Cys	Leu	Val	Leu	Ser	Leu	Asp	Asn	Trp	100	105	110	
Ser	Val	Trp	Cys	Tyr	Val	Cys	Asp	Asn	Glu	Val	Gln	Tyr	Cys	Ser	Ser	115	120	125	
Asn	Gln	Leu	Gly	Gln	Val	Val	Asp	Tyr	Val	Arg	Lys	Gln	Ala	Ser	Ile	130	135	140	
Thr	Thr	Pro	Lys	Pro	Ala	Glu	Lys	Asp	Asn	Gly	Asn	Ile	Glu	Leu	Glu	145	150	155	160
Asn	Lys	Lys	Leu	Glu	Lys	Glu	Ser	Lys	Asn	Glu	Gln	Glu	Arg	Glu	Lys	165	170	175	
Lys	Glu	Asn	Met	Ala	Lys	Glu	Asn	Pro	Pro	Met	Asn	Ser	Pro	Cys	Gln	180	185	190	
Ile	Thr	Val	Lys	Gly	Leu	Ser	Asn	Leu	Gly	Asn	Thr	Cys	Phe	Phe	Asn	195	200	205	
Ala	Val	Met	Gln	Asn	Leu	Ser	Gln	Thr	Pro	Val	Leu	Arg	Glu	Leu	Leu	210	215	220	
Lys	Glu	Val	Lys	Met	Ser	Gly	Thr	Ile	Val	Lys	Ile	Glu	Pro	Pro	Asp	225	230	235	240
Leu	Ala	Leu	Thr	Glu	Pro	Leu	Glu	Ile	Asn	Leu	Glu	Pro	Pro	Gly	Pro	245	250	255	
Leu	Thr	Leu	Ala	Met	Ser	Gln	Phe	Leu	Asn	Glu	Met	Gln	Glu	Thr	Lys	260	265	270	
Lys	Gly	Val	Val	Thr	Pro	Lys	Glu	Leu	Phe	Ser	Gln	Val	Cys	Lys	Lys	275	280	285	
Ala	Val	Arg	Phe	Lys	Gly	Tyr	Gln	Gln	Gln	Asp	Ser	Gln	Glu	Leu	Leu	290	295	300	
Arg	Tyr	Leu	Leu	Asp	Gly	Met	Arg	Ala	Glu	Glu	His	Gln	Arg	Val	Ser	305	310	315	320
Lys	Gly	Ile	Leu	Lys	Ala	Phe	Gly	Asn	Ser	Thr	Glu	Lys	Leu	Asp	Glu	325	330	335	
Glu	Leu	Lys	Asn	Lys	Val	Lys	Asp	Tyr	Glu	Lys	Lys	Lys	Ser	Met	Pro	340	345	350	
Ser	Phe	Val	Asp	Arg	Ile	Phe	Gly	Gly	Glu	Leu	Thr	Ser	Met	Ile	Met	355	360	365	
Cys	Asp	Gln	Cys	Arg	Thr	Val	Ser	Leu	Val	His	Glu	Ser	Phe	Leu	Asp	370	375	380	
Leu	Ser	Leu	Pro	Val	Leu	Asp	Asp	Gln	Ser	Gly	Lys	Lys	Ser	Val	Asn	385	390	395	400
Asp	Lys	Asn	Leu	Lys	Lys	Thr	Val	Glu	Asp	Glu	Asp	Gln	Asp	Ser	Glu	405	410	415	
Glu	Glu	Lys	Asp	Asn	Asp	Ser	Tyr	Ile	Lys	Glu	Arg	Ser	Asp	Ile	Pro	420	425	430	
Ser	Gly	Thr	Ser	Lys	His	Leu	Gln	Lys	Lys	Ala	Lys	Lys	Gln	Ala	Lys	435	440	445	
Lys	Gln	Ala	Lys	Asn	Gln	Arg	Arg	Gln	Gln	Lys	Ile	Gln	Gly	Lys	Val	450	455	460	
Leu	His	Leu	Asn	Asp	Ile	Cys	Thr	Ile	Asp	His	Pro	Glu	Asp	Ser	Glu	465	470	475	480
Tyr	Glu	Ala	Glu	Met	Ser	Leu	Gln	Gly	Glu	Val	Asn	Ile	Lys	Ser	Asn				

	485		490		495
His Ile Ser	Gln Glu Gly Val Met	His Lys Glu Tyr Cys Val	Asn Gln		
	500	505	510		
Lys Asp Leu	Asn Gly Gln Ala Lys	Met Ile Glu Ser Val Thr	Asp Asn		
	515	520	525		
Gln Lys Ser	Thr Glu Glu Val Asp	Met Lys Asn Ile Asn Met	Asp Asn		
	530	535	540		
Asp Leu Glu	Val Leu Thr Ser Ser	Pro Thr Arg Asn Leu Asn	Gly Ala		
	545	550	555	560	
Tyr Leu Thr	Glu Gly Ser Asn Gly	Glu Val Asp Ile Ser Asn	Gly Phe		
	565	570	575		
Lys Asn Leu	Asn Leu Asn Ala Ala	Leu His Pro Asp Glu Ile	Asn Ile		
	580	585	590		
Glu Ile Leu	Asn Asp Ser His Thr	Pro Gly Thr Lys Val Tyr	Glu Val		
	595	600	605		
Val Asn Glu	Asp Pro Glu Thr Ala	Phe Cys Thr Leu Ala	Asn Arg Glu		
	610	615	620		
Val Phe Asn	Thr Asp Glu Cys Ser	Ile Gln His Cys Leu Tyr	Gln Phe		
	625	630	635	640	
Thr Arg Asn	Glu Lys Leu Arg Asp	Ala Asn Lys Leu Leu Cys	Glu Val		
	645	650	655		
Cys Thr Arg	Arg Gln Cys Asn Gly	Pro Lys Ala Asn Ile Lys	Gly Glu		
	660	665	670		
Arg Lys His	Val Tyr Thr Asn Ala	Lys Lys Gln Met Leu Ile	Ser Leu		
	675	680	685		
Ala Pro Pro	Val Leu Thr Leu His	Leu Lys Arg Phe Gln	Gln Ala Gly		
	690	695	700		
Phe Asn Leu	Arg Lys Val Asn Lys	His Ile Lys Phe Pro	Glu Ile Leu		
	705	710	715	720	
Asp Leu Ala	Pro Phe Cys Thr	Leu Lys Cys Lys Asn	Val Ala Glu	Glu	
	725	730	735		
Asn Thr Arg	Val Leu Tyr Ser	Leu Tyr Gly Val Val	Glu His Ser	Gly	
	740	745	750		
Thr Met Arg	Ser Gly His Tyr	Thr Ala Tyr Ala Lys	Ala Arg Thr	Ala	
	755	760	765		
Asn Ser His	Leu Ser Asn Leu	Val Leu His Gly Asp	Ile Pro Gln	Asp	
	770	775	780		
Phe Glu Met	Glu Ser Lys Gly	Gln Trp Phe His	Ile Ser Asp	Thr His	
	785	790	795	800	
Val Gln Ala	Val Pro Thr Thr	Lys Val Leu Asn Ser	Gln Ala Tyr	Leu	
	805	810	815		
Leu Phe Tyr	Glu Arg Ile Leu				
	820				

<210> 402  
 <211> 2903  
 <212> DNA  
 <213> Homo sapiens

<400> 402  
 ctcaattcgt caccaggagg aagacggagc tggctgccca gcccaaaggc ccatgagggg 60  
 atgcagttat gggctctgtc gccgtggatt gttattttgt gtcagtaagt aatccataaa 120  
 gtgccaacat gggaaagaaa cggacaaagg gaaaaactgt tccaatcgat gattcctctg 180  
 aaactttaga acctgtgtgc agacacatta gaaaaggatt ggaacaaggt aatttgaaaa 240  
 aggctttagt gaatgtggaa tggaatatct gccaaactgt taagactgac aataaagtga 300  
 aagataaagc tgaagaagaa aattctcagg agccttcagt ttggctgtgt cttaaatgtg 360  
 gccatcaggg ctgtggcaga aattctcagg agcagcatgc cttgaagcac tatctgacgc 420  
 caagatctga acctcactgt ctggttctta gtttggacaa ctggagtgtg tggtgttacg 480  
 tatgtgataa tgaggtccag tattgtagtt caaaccagtt gggccaagt gttgattatg 540  
 tcagaaaaca agccagcatt acaactccaa agccagcaga gaaagataat ggaaatattg 600

```

aacttgaaaa taaaaaatta gaaaaagaga gtaagaatga acaagagaga gaaaagaagg 660
aaaacatggc taaagagaat cctcccatga attctccttg ccaaataacc gtgaaaggac 720
tcagtaattt gggaaacaca tgtttcttca atgcagttat gcagaacttg tcacaaacac 780
cagtgttag agaaactacta aaagaagtga aaatgtctgg aacaattgta aaaattgaac 840
cacctgattt ggcattaaca gaaccattag aaataaacct tgagcctcca ggccctctta 900
cttttagccat gagccagttt cttaatgaga tgcaagagac caaaaagggg gttgtgacac 960
cgaaagaact cttttctcag gtctgtaaaa aagcagtgcg gtttaaaggc tatcagcagc 1020
aagacagcca ggaagctgctt cgctacttat tggatgggat gagagcagaa gaacaccaaa 1080
gagttagtaa aggaatactt aaagcatttg gtaattctac tgaaaagtgt gatgaagaac 1140
taaaaaataa agttaagat tatgagaaga aaaaatcaat gccaaagttt gttgaccgca 1200
tctttgtgg tgaactaact agtatgatca tgtgtgatca atgcagaact gtctccttgg 1260
ttcatgaatc tttccttgat ttgtccctcc cagttttaga tgatcagagt ggtaagaaaa 1320
gtgtaaatga taaaaatctg aaaaagacag tggaggatga agatcaagat agtgaggaag 1380
aaaaagataa cgacagttac ataaaagaga gaagtgatat tccttctgga acaagtaagc 1440
acttacagaa aaaagcaaaag aaacaagcca aaaagcaagc caagaacca cgaagacaa 1500
aaaaaattca aggaaaagtt cttcatttaa atgatatttg tactattgac catcctgaag 1560
acagtgaata tgaagctgaa atgtcacttc aaggagaagt aaatattaaa tccaaccata 1620
tttcacaaga ggtgttatg cataaagaat attgtgtcaa ccagaaagat ttgaatggcc 1680
aagcaaaaat gatcgaaagt gtaactgaca atcaaaaatc cacagaggaa gtagatatga 1740
aaaatatcaa catggataat gatctggagg ttttaacatc ttctcccact aggaatttaa 1800
atggtgccta cctaacggaa gggagcaatg gagaagtga catttccaat ggtttcaaaa 1860
acctaaattt gaatgtgct cttcatcctg atgaaataaa tatagagatt ctgaatgata 1920
gtcactactcc tggaaacaaag gtgtatgagg ttgtaaatga agatccagaa actgctttct 1980
gtactcttgc aaacagggaa gttttcaata ctgatgagt ttcaatccaa cattgtttat 2040
atcagttcac ccgtaatgag aaacttcgag atgcgaataa actgctttgt gaagtatgca 2100
cacggagaca gtgtaatgga ccaaaggcaa atataaaaagg tgaaagggaag catgtttaca 2160
ccaatgccaa aaagcagatg ctaatttctc ttgtcctcc tgttcttact cttcatttaa 2220
agagatttca gcagctggg tttaacctac gcaaagttaa caaacacata aagtttccgg 2280
aaatcttaga tttggctcct ttttgaccc ttaaattgaa gaatgttga gaagaaaata 2340
caagggtagt ctattcctta tatggagttg ttgaacacag tggtagtagt aggtcggggc 2400
attactctgc ctatgccaa gcaagaaccg caaatagtca tctctcta at cttgttcttc 2460
acggtgatat tccacaagat tttgaaatgg aatcaaaaagg gcagtggttt cacatcagcg 2520
acacacatgt gcaagctgtg cctacaacta aagtactaaa ctcacaagcg tacctcctat 2580
tttatgagag aatactgtaa taatatcaaa agcacttttt ctggaaacac atttatggct 2640
tttataatgg ctgaaataac gataaaaaaa gactaattaa aatcatgttc acttaacct 2700
aaatacatgc cagaagaaat catgtttatt taaatattga agggaaaaat acctaaaaat 2760
gtacaaaagg tttatattgt catagtgggt tttattcctg ctttgtttct ggaaaggaaa 2820
tctgaatta cttaagtact ttgtgtttaa tatatctggg tgatggatca caacacatca 2880
ataaactgac ttaccctaaa atc 2903

```

<210> 403  
 <211> 205  
 <212> PRT  
 <213> Homo sapiens

<400> 403  
 Met Ala Leu Gln Leu Ser Arg Glu Gln Gly Ile Thr Leu Arg Gly Ser  
 1 5 10 15  
 Ala Glu Ile Val Ala Glu Phe Phe Ser Phe Gly Ile Asn Ser Ile Leu  
 20 25 30  
 Tyr Gln Arg Gly Ile Tyr Pro Ser Glu Thr Phe Thr Arg Val Gln Lys  
 35 40 45  
 Tyr Gly Leu Thr Leu Leu Val Thr Thr Asp Leu Glu Leu Ile Lys Tyr  
 50 55 60  
 Leu Asn Asn Val Val Glu Gln Leu Lys Asp Trp Leu Tyr Lys Cys Ser  
 65 70 75 80  
 Val Gln Lys Leu Val Val Val Ile Ser Asn Ile Glu Ser Gly Glu Val  
 85 90 95  
 Leu Glu Arg Trp Gln Phe Asp Ile Glu Cys Asp Lys Thr Ala Lys Asp  
 100 105 110  
 Asp Ser Ala Pro Arg Glu Lys Ser Gln Lys Ala Ile Gln Asp Glu Ile

```

      115      120      125
Arg Ser Val Ile Arg Gln Ile Thr Ala Thr Val Thr Phe Leu Pro Leu
      130      135      140
Leu Glu Val Ser Cys Ser Phe Asp Leu Leu Ile Tyr Thr Asp Lys Asp
      145      150      155      160
Leu Val Val Pro Glu Lys Trp Glu Glu Ser Gly Pro Gln Phe Ile Thr
      165      170      175
Asn Ser Glu Glu Val Arg Leu Arg Ser Phe Thr Thr Thr Ile His Lys
      180      185      190
Val Asn Ser Met Val Ala Tyr Lys Ile Pro Val Asn Asp
      195      200      205

```

<210> 404  
 <211> 1390  
 <212> DNA  
 <213> Homo sapiens

```

<400> 404
gggaagtgct gttggagccg ctgtggttgc tgtccgcgga gtggaagcgc gtgcttttgt 60
ttgtgtccct ggccatggcg ctgcagctct cccgggagca gggaatcacc ctgcgcggga 120
gcgccgaaat cgtggccgag ttcttctcat tcggcatcaa cagcatttta tatcagcgtg 180
gcataatatc atctgaaacc tttactcgag tgcagaaata cggactcacc ttgcttgtaa 240
ctactgatct tgagctcata aaatacctaa ataatgtggt ggaacaactg aaagattggt 300
tatacaagtg ttcagttcag aaactgggtg tagttatctc aaatatgaa agtggtgagg 360
tcctggaaag atggcagttt gatattgagt gtgacaagac tgcaaaagat gacagtgcac 420
ccagagaaaa gtctcagaaa gctatccagg atgaaatccg ttcagtgatc agacagatca 480
cagctacggt gacatttctg ccactgttgg aagtttcttg ttcatttgat ctgctgattt 540
atacagacaa agatttggtt gtacctgaaa aatgggaaga gtcgggacca cagtttatta 600
ccaattctga ggaagtccgc cttcgttcat ttactactac aatccacaaa gtaaatagca 660
tggtggccta caaaattcct gtcaatgact gaggatgaca tgaggaaaat aatgtaattg 720
taattttgaa atgtgggttt cctgaaatca atgatactta ctgaactgtg tgtaattggt 780
ttggttaatt tttacatgga gaaaaccaa ccatggagtt aacatcatga atttattgca 840
cctttatttt tttggtacct atttgactta ccatggagtt aacatcatga atttattgca 900
cattgttcaa aaggaaccag gaggtttttt tgtcaacatt gtgatgtata ttcctttgaa 960
gatagtaact gtagatggaa aaacttgtgc tataaagcta gatgctttcc taaatcagat 1020
gttttggtca agtagtttga ctcagtatag gtagggagat atttaagtat aaaatacaac 1080
aaaggaagtc taaatattca gaatccttgt taaggtcctg aaagtaactc ataactctata 1140
aacaatgaaa tattgtctga tagctccttt tgaccttcat ttcattgtata gttttcccta 1200
ttgaatcagt ttccaattat ttgactttta tttatgtaac ttgaacctat gaagcaatgg 1260
atatttgtac tgtttaatgt tctgtgatac agaactctta aaaatgtttt ttcattgtgt 1320
ttataaaatc aagtttttaag tgaaagtggag gaaataaagt taagtttggt ttaaaaaaaaa 1380
aaaaaaaaaa                                     1390

```

<210> 405  
 <211> 464  
 <212> PRT  
 <213> Homo sapiens

```

<400> 405
Met Glu Thr Leu Ser Phe Pro Arg Tyr Asn Val Ala Glu Ile Val Ile
 1      5      10      15
His Ile Arg Asn Lys Ile Leu Thr Gly Ala Asp Gly Lys Asn Leu Thr
      20      25      30
Lys Asn Asp Leu Tyr Pro Asn Pro Lys Pro Glu Val Leu His Met Ile
      35      40      45
Tyr Met Arg Ala Leu Gln Ile Val Tyr Gly Ile Arg Leu Glu His Phe
      50      55      60
Tyr Met Met Pro Val Asn Ser Glu Val Met Tyr Pro His Leu Met Glu
      65      70      75      80
Gly Phe Leu Pro Phe Ser Asn Leu Val Thr His Leu Asp Ser Phe Leu

```

```
<210> 406
<211> 1857
<212> DNA
<213> Homo sapiens
```

-168-

```

actctgaagt catgtatcca catttaatgg aaggcttctt accattcagc aatttagtta 420
ctcatctgga ctcatcttctg cctatctgcc ggggtgaatga ctttgagact gctgatattc 480
tatgtccaaa agcaaaacgg acaagtcggt ttttaagtgg cattatcaac tttattcact 540
tcagagaagc atgccgtgaa acgtatatgg aatttctttg gcaatataaa tcctctgctg 600
acaaaatgca acagttaaac gccgcacacc aggaggcatt aatgaaactg gagagacttg 660
attctgttcc agttgaagag caagaagagt tcaagcagct ttcagatggg attcaggagc 720
tacaacaatc actaaatcag gattttcatt aaaaaacgat agtgctgcaa gagggaaatt 780
cccaaaagaa gtcaaatatt tcagagaaaa ccaagcggtt gaatgaacta aaattgttgg 840
tggtttcttt gaaagaaata caagagagtt tgaaaaacaa aattgtggat tctccagaga 900
agttaaagaa ttataaagaa aaaatgaaag atacgggtcca gaagcttaaa aatgccagac 960
aagaagtggg ggagaaatat gaaatctatg gagactcagt tgactgcctg ccttcatgtc 1020
agttggaagt gcagttatat caaaagaaaa tacaggacct ttcagataat agggaaaaat 1080
tagccagtat cttaaaggag agcctgaact tggaggacca aattgagagt gatgagtcag 1140
aactgaagaa attgaagact gaagaaaatt cgttcaaaag actgatgatt gtgaagaagg 1200
aaaaacttgc cacagcacaa ttcaaaataa ataagaagca tgaagatgtt aagcaatata 1260
aacgcacagt aattgaggat tgcaataaag ttcaagaaaa aagagggtgt gtctatgaac 1320
gagtaaccac aattaatcaa gaaatccaaa aaattaaact tggaaattcaa caactaaaag 1380
atgctgctga aaggagaaaa ctgaagtccc aggaatatatt tctaaacttg aaaactgctt 1440
tgagagaaata ccacgacggt attgaaaagg cagcagagga ctctatgct agatagatg 1500
agaagacagc tgaactgaag aggaagatgt tcaaaatgtc aacctgatta acaaaattac 1560
atgtcttttt gtaaatggct tgccatcttt taattttcta tttagaaaga aaagttgaag 1620
cgaatggaag tatcagaagt accaaataat gttggcttca tcagttttta tacactctca 1680
taagtagtta ataagatgaa tttaatgtag gcttttatta atttataatt aaaataactt 1740
gtgcagctat tcatgtctct actctgcccc ttggtgtaaa tagtttgagt aaaacaaaac 1800
tagttacctt tgaaatatat atattttttt ctgttaaaaa aaaaaaaa 1857

```

&lt;210&gt; 407

&lt;211&gt; 1050

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 407

```

Met Ala Ala Val Lys Lys Glu Gly Gly Ala Leu Ser Glu Ala Met Ser
1      5      10      15
Leu Glu Gly Asp Glu Trp Glu Leu Ser Lys Glu Asn Val Gln Pro Leu
20      25      30
Arg Gln Gly Arg Ile Met Ser Thr Leu Gln Gly Ala Leu Ala Gln Glu
35      40      45
Ser Ala Cys Asn Asn Thr Leu Gln Gln Gln Lys Arg Ala Phe Glu Tyr
50      55      60
Glu Ile Arg Phe Tyr Thr Gly Asn Asp Pro Leu Asp Val Trp Asp Arg
65      70      75      80
Tyr Ile Ser Trp Thr Glu Gln Asn Tyr Pro Gln Gly Gly Lys Glu Ser
85      90      95
Asn Met Ser Thr Leu Leu Glu Arg Ala Val Glu Ala Leu Gln Gly Glu
100     105     110
Lys Arg Tyr Tyr Ser Asp Pro Arg Phe Leu Asn Leu Trp Leu Lys Leu
115     120     125
Gly Arg Leu Cys Asn Glu Pro Leu Asp Met Tyr Ser Tyr Leu His Asn
130     135     140
Gln Gly Ile Gly Val Ser Leu Ala Gln Phe Tyr Ile Ser Trp Ala Glu
145     150     155     160
Glu Tyr Glu Ala Arg Glu Asn Phe Arg Lys Ala Asp Ala Ile Phe Gln
165     170     175
Glu Gly Ile Gln Lys Ala Glu Pro Leu Glu Arg Leu Gln Ser Gln
180     185     190
His Arg Gln Phe Gln Ala Arg Val Ser Arg Gln Thr Leu Leu Ala Leu
195     200     205
Glu Lys Glu Glu Glu Glu Glu Val Phe Glu Ser Ser Val Pro Gln Arg
210     215     220
Ser Thr Leu Ala Glu Leu Lys Ser Lys Gly Lys Lys Thr Ala Arg Ala

```

225					230					235					240
Pro	Ile	Ile	Arg	Val	Gly	Gly	Ala	Leu	Lys	Ala	Pro	Ser	Gln	Asn	Arg
				245					250					255	
Gly	Leu	Gln	Asn	Pro	Phe	Pro	Gln	Gln	Met	Gln	Asn	Asn	Ser	Arg	Ile
			260					265					270		
Thr	Val	Phe	Asp	Glu	Asn	Ala	Asp	Glu	Ala	Ser	Thr	Ala	Glu	Leu	Ser
		275					280					285			
Lys	Pro	Thr	Val	Gln	Pro	Trp	Ile	Ala	Pro	Pro	Met	Pro	Arg	Ala	Lys
	290					295					300				
Glu	Asn	Glu	Leu	Gln	Ala	Gly	Pro	Trp	Asn	Thr	Gly	Arg	Ser	Leu	Glu
305					310					315				320	
His	Arg	Pro	Arg	Gly	Asn	Thr	Ala	Ser	Leu	Ile	Ala	Val	Pro	Ala	Val
				325					330					335	
Leu	Pro	Ser	Phe	Thr	Pro	Tyr	Val	Glu	Glu	Thr	Ala	Gln	Gln	Pro	Val
			340					345				350			
Met	Thr	Pro	Cys	Lys	Ile	Glu	Pro	Ser	Ile	Asn	His	Ile	Leu	Ser	Thr
		355					360					365			
Arg	Lys	Pro	Gly	Lys	Glu	Glu	Gly	Asp	Ser	Leu	Gln	Arg	Val	Gln	Ser
	370					375					380				
His	Gln	Gln	Ala	Ser	Glu	Glu	Lys	Lys	Glu	Lys	Met	Met	Tyr	Cys	Lys
385					390					395				400	
Glu	Lys	Ile	Tyr	Ala	Gly	Val	Gly	Glu	Phe	Ser	Phe	Glu	Glu	Ile	Arg
			405					410						415	
Ala	Glu	Val	Phe	Arg	Lys	Lys	Leu	Lys	Glu	Gln	Arg	Glu	Ala	Glu	Leu
		420					425					430			
Leu	Thr	Ser	Ala	Glu	Lys	Arg	Ala	Glu	Met	Gln	Lys	Gln	Ile	Glu	Glu
	435						440					445			
Met	Glu	Lys	Lys	Leu	Lys	Glu	Ile	Gln	Thr	Thr	Gln	Gln	Glu	Arg	Thr
	450					455					460				
Gly	Asp	Gln	Gln	Glu	Glu	Thr	Met	Pro	Thr	Lys	Glu	Thr	Thr	Lys	Leu
465					470					475				480	
Gln	Ile	Ala	Ser	Glu	Ser	Gln	Lys	Ile	Pro	Gly	Met	Thr	Leu	Ser	Ser
			485						490					495	
Ser	Val	Cys	Gln	Val	Asn	Cys	Cys	Ala	Arg	Glu	Thr	Ser	Leu	Ala	Glu
		500						505				510			
Asn	Ile	Trp	Gln	Glu	Gln	Pro	His	Ser	Lys	Gly	Pro	Ser	Val	Pro	Phe
	515						520					525			
Ser	Ile	Phe	Asp	Glu	Phe	Leu	Leu	Ser	Glu	Lys	Lys	Asn	Lys	Ser	Pro
	530					535					540				
Pro	Ala	Asp	Pro	Pro	Arg	Val	Leu	Ala	Gln	Arg	Arg	Pro	Leu	Ala	Val
545					550					555				560	
Leu	Lys	Thr	Ser	Glu	Ser	Ile	Thr	Ser	Asn	Glu	Asp	Val	Ser	Pro	Asp
			565						570					575	
Val	Cys	Asp	Glu	Phe	Thr	Gly	Ile	Glu	Pro	Leu	Ser	Glu	Asp	Ala	Ile
		580						585					590		
Ile	Thr	Gly	Phe	Arg	Asn	Val	Thr	Ile	Cys	Pro	Asn	Pro	Glu	Asp	Thr
	595						600					605			
Cys	Asp	Phe	Ala	Arg	Ala	Ala	Arg	Phe	Val	Ser	Thr	Pro	Phe	His	Glu
	610					615					620				
Ile	Met	Ser	Leu	Lys	Asp	Leu	Pro	Ser	Asp	Pro	Glu	Arg	Leu	Leu	Pro
625					630					635				640	
Glu	Glu	Asp	Leu	Asp	Val	Lys	Thr	Ser	Glu	Asp	Gln	Gln	Thr	Ala	Cys
			645						650					655	
Gly	Thr	Ile	Tyr	Ser	Gln	Thr	Leu	Ser	Ile	Lys	Lys	Leu	Ser	Pro	Ile
		660						665					670		
Ile	Glu	Asp	Ser	Arg	Glu	Ala	Thr	His	Ser	Ser	Gly	Phe	Ser	Gly	Ser
	675						680					685			
Ser	Ala	Ser	Val	Ala	Ser	Thr	Ser	Ser	Ile	Lys	Cys	Leu	Gln	Ile	Pro
	690					695					700				
Glu	Lys	Leu	Glu	Leu	Thr	Asn	Glu	Thr	Ser	Glu	Asn	Pro	Thr	Gln	Ser
705					710					715				720	

Pro Trp Cys Ser Gln Tyr Arg Arg Gln Leu Leu Lys Ser Leu Pro Glu  
 725 730 735  
 Leu Ser Ala Ser Ala Glu Leu Cys Ile Glu Asp Arg Pro Met Pro Lys  
 740 745 750  
 Leu Glu Ile Glu Lys Glu Ile Glu Leu Gly Asn Glu Asp Tyr Cys Ile  
 755 760 765  
 Lys Arg Glu Tyr Leu Ile Cys Glu Asp Tyr Lys Leu Phe Trp Val Ala  
 770 775 780  
 Pro Arg Asn Ser Ala Glu Leu Thr Val Ile Lys Val Ser Ser Gln Pro  
 785 790 795 800  
 Val Pro Trp Asp Phe Tyr Ile Asn Leu Lys Leu Lys Glu Arg Leu Asn  
 805 810 815  
 Glu Asp Phe Asp His Phe Cys Ser Cys Tyr Gln Tyr Gln Asp Gly Cys  
 820 825 830  
 Ile Val Trp His Gln Tyr Ile Asn Cys Phe Thr Leu Gln Asp Leu Leu  
 835 840 845  
 Gln His Ser Glu Tyr Ile Thr His Glu Ile Thr Val Leu Ile Ile Tyr  
 850 855 860  
 Asn Leu Leu Thr Ile Val Glu Met Leu His Lys Ala Glu Ile Val His  
 865 870 875 880  
 Gly Asp Leu Ser Pro Arg Cys Leu Ile Leu Arg Asn Arg Ile His Asp  
 885 890 895  
 Pro Tyr Asp Cys Asn Lys Asn Asn Gln Ala Leu Lys Ile Val Asp Phe  
 900 905 910  
 Ser Tyr Ser Val Asp Leu Arg Val Gln Leu Asp Val Phe Thr Leu Ser  
 915 920 925  
 Gly Phe Arg Thr Val Gln Ile Leu Glu Gly Gln Lys Ile Leu Ala Asn  
 930 935 940  
 Cys Ser Ser Pro Tyr Gln Val Asp Leu Phe Gly Ile Ala Asp Leu Ala  
 945 950 955 960  
 His Leu Leu Leu Phe Lys Glu His Leu Gln Val Phe Trp Asp Gly Ser  
 965 970 975  
 Phe Trp Lys Leu Ser Gln Asn Ile Ser Glu Leu Lys Asp Gly Glu Leu  
 980 985 990  
 Trp Asn Lys Phe Phe Val Arg Ile Leu Asn Ala Asn Asp Glu Ala Thr  
 995 1000 1005  
 Val Ser Val Leu Gly Glu Leu Ala Ala Glu Met Asn Gly Val Phe Asp  
 1010 1015 1020  
 Thr Thr Phe Gln Ser His Leu Asn Lys Ala Leu Trp Lys Val Gly Lys  
 1025 1030 1035 1040  
 Leu Thr Ser Pro Gly Ala Leu Leu Phe Gln  
 1045 1050

<210> 408  
 <211> 3702  
 <212> DNA  
 <213> Homo sapiens

<400> 408  
 gtttgtagg gagtcgtgtg cgtgccttgg tcgcttctgt agctccgagg gcaggttgcg 60  
 gaagaaagcc caggcgggtct gtggcccaga ggaaaggcct gcagcaggac gaggacctga 120  
 gccaggaatg caggatggcg gcggtgaaga aggaaggggg tgctctgagt gaagccatgt 180  
 ccctggaggg agatgaatgg gaactgagta aagaaaatgt acaaccttta aggcaagggc 240  
 ggatcatgtc cagccttcag ggagcactgg cacaagaatc tgccctgtaac aatactcttc 300  
 agcagcagaa acgggcattt gaatatgaaa ttcgatttta cactggaaat gaccctcttg 360  
 atgtttggga taggtatata agctggacag agcagaacta tcctcaaggt gggaaggaga 420  
 gtaatatgtc aacgttatta gaaagagctg tagaagcact acaaggagaa aaacgatatt 480  
 atagtgtacc tcgatttctc aatctctggc ttaaattagg gcgtttatgc aatgagcctt 540  
 tggtatgtga cagttacttg cacaaccaag ggattggtgt ttcacttgct cagttctata 600  
 tctcatgggc agaagaatat gaagctagag aaaactttag gaaagcagat gcgatatttc 660



```

aggaagggat tcaacagaag gctgaaccac tagaaagact acagtcccag caccgacaat 720
tccaagctcg agtgtctcgg caaactctgt tggcacttga gaaagaagaa gaggagggaag 780
tttttgagtc ttctgtacca caacgaagca cactagctga actaaagagc aaagggaaaa 840
agacagcaag agctccaatc atccgtgtag gaggtgctct caaggctcca agccagaaca 900
gaggactcca aaatccattt cctcaacaga tgcaaaataa tagtagaatt actgtttttg 960
atgaaaatgc tgatgaggct tctacagcag agttgtctaa gcctacagtc cagccatgga 1020
tagcaccccc catgcccagg gccaaagaga atgagctgca agcaggccct tggaaacacag 1080
gcaggtcctt ggaacacagg cctcgtggca atacagcttc actgatagct gtaccgctg 1140
tgcttcccag tttactcca tatgtggaag agactgcaca acagccagtt atgacaccat 1200
gtaaaattga acctagtata aaccacatcc taagcaccag aaagcctgga aaggaagaag 1260
gagattctct acaaagggtt cagagccatc agcaagcgtc tgaggagaag aaagagaaga 1320
tgatgtattg taaggagaag atttatgcag gagtagggga attctccttt gaagaaattc 1380
gggctgaagt tttccggaag aaattaaaag agcaaaggga agccgagcta ttgaccagt 1440
cagagaagag atcagaaatg cagaaacaga ttgaagagat ggagaagaag ctaaaagaaa 1500
tccaaactac tcagcaagaa agaacaggta atcagcaaga agagacgatg cctacaaagg 1560
agacaactaa actgcaaatt gcttccgagt ctcagaaaaa accaggaatg actctatcca 1620
gttctgtttg tcaagtaaac tgttgtgcca gagaaacttc acttgccggag aacatttggc 1680
aggaacaacc tcattctaaa ggtcccagtg tacctttctc catttttgat gagtttcttc 1740
tttcagaaaa gaagaataaa agtcctcctg cagatcccc acgagtttta gctcaacgaa 1800
gaccccttgc agttctcaaa acctcagaaa gcatacctc aaatgaagat gtgtctccag 1860
atgtttgtga tgaatttaca ggaattgaac ccttgagcga ggatgccatt atcacaggct 1920
tcagaaatgt aacaatttgt cctaaccag aagacacttg tgactttgcc agagcagctc 1980
gttttgtatc cactcctttt catgagataa tgtccttgaa ggatctccct tctgatcctg 2040
agagactggt accggaagaa gatctagatg taaagacctc tgaggaccag cagacagctt 2100
gtggcactat ctacagtcag actctcagca tcaagaagct gagcccaatt attgaagaca 2160
gtcgtgaagc cacacactcc tctggcttct ctggttcttc tgcctcggtt gcaagcacct 2220
cctccatcaa atgtcttcaa attcctgaga aactagaact tactaatgag acttcagaaa 2280
accctactca gtcaccattg tgttcacagt atcgagaca gctactgaag tccctaccag 2340
agttaagtgc ctctgcagag ttgtgtatag aagacagacc aatgcctaag ttggaattg 2400
agaaggaaat tgaattaggt aatgaggatt actgcattaa acgagaatac ctaatatgtg 2460
aagattacaa gttattctgg gtggcgccaa gaaactctgc agaattaaca gtaataaagg 2520
tatcttctca acctgtccca tgggactttt atatcaacct caagttaaag gaacgtttaa 2580
atgaagattt tgatcatttt tgcagctggt atcaatatca agatggctgt attgtttggc 2640
accaatatat aaactgcttc accttcagg atcttctcca acacagtga tatattaccc 2700
atgaaataac agtgttgatt atttataacc ttttgacaat agtggagatg ctacacaaag 2760
cagaaatagt ccatggtgac ttgagtccaa ggtgtctgat tctcagaaac agaatccacg 2820
atccctatga ttgtaacaag aacaatcaag ctttgaagat agtggacttt tcctacagt 2880
ttgaccttag ggtgcagctg gatgttttta cctcagcgg ctttcggact gtacagatcc 2940
tggaaggaca aaagatcctg gctaactgtt cttctcccta ccaggtagac ctgtttggta 3000
tagcagattt agcacattta ctattgttca aggaacacct acaggtcttc tgggatgggt 3060
ccttctggaa acttagccaa aatatttctg agctaaaaga tggatgaattg tggataaaat 3120
tctttgtgcg gatttggaat gccaatgatg aggccacagt gtctgttctt ggggagcttg 3180
cagcagaaat gaatggggtt tttgacacta cattccaaag tcacctgaac aaagccttat 3240
ggaaggtagg gaagttaact agtcctgggg ctttgcctt tcagtgaagt aggcaatcaa 3300
gtctcacaga ttgctgcctc agagcaatgg ttgtattgtg gaacactgaa actgtatgtg 3360
ctgtaattta atttaggaca catttagatg cactaccatt gctgttctac tttttgggtac 3420
aggtatatatt tgacgtcact gatatttttt atacagtgat atacttactc atggccttgt 3480
ctaacttttg tgaagaacta ttttattcta aacagactca ttacaaatgg ttaccttgtt 3540
atttaaccca tttgtctcta ctttccctg tacttttccc atttgtaatt tgtaaaatgt 3600
tctcttatga tcaccatgta ttttgtaaat aataaaatag tatctgttaa aaaaaaaaaa 3660
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aa 3702

```

<210> 409  
 <211> 164  
 <212> PRT  
 <213> Homo sapiens

<400> 409  
 Met Pro Cys Ser Glu Glu Thr Pro Ala Ile Ser Pro Ser Lys Arg Ala  
 1 5 10 15  
 Arg Pro Ala Glu Val Gly Gly Met Gln Leu Arg Phe Ala Arg Leu Ser

20 25 30  
 Glu His Ala Thr Ala Pro Thr Arg Gly Ser Ala Arg Ala Gly Tyr  
 35 40 45  
 Asp Leu Tyr Ser Ala Tyr Asp Tyr Thr Ile Pro Pro Met Glu Lys Ala  
 50 55 60  
 Val Val Lys Thr Asp Ile Gln Ile Ala Leu Pro Ser Gly Cys Tyr Gly  
 65 70 75 80  
 Arg Val Ala Pro Arg Ser Gly Leu Ala Ala Lys His Phe Ile Asp Val  
 85 90 95  
 Gly Ala Gly Val Ile Asp Glu Asp Tyr Arg Gly Asn Val Gly Val Val  
 100 105 110  
 Leu Phe Asn Phe Gly Lys Glu Lys Phe Glu Val Lys Lys Gly Asp Arg  
 115 120 125  
 Ile Ala Gln Leu Ile Cys Glu Arg Ile Phe Tyr Pro Glu Ile Glu Glu  
 130 135 140  
 Val Gln Ala Leu Asp Asp Thr Glu Arg Gly Ser Gly Gly Phe Gly Ser  
 145 150 155 160  
 Thr Gly Lys Asn

&lt;210&gt; 410

&lt;211&gt; 1816

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 410

ctgcgcttct ggctctgcca tgccctgctc tgaagagaca cccgccattt caccagtaa 60  
 gcgggcccgg cctgcggagg tgggcggcat gcagctccgc tttgcccggc tctccgagca 120  
 cgccacggcc cccaccggg gctccgcgcg cgccgcgggc tacgacctgt acagtgccta 180  
 tgattacaca ataccaccta tggagaaagc tgttgtagaa acggacattc agatagcgct 240  
 cccttctggg tggtatggaa gagtggctcc acggtcaggc ttggctgcaa aacactttat 300  
 tgatgtagga gctgggtgta tagatgaaga ttatagagga aatgttggtg ttgtactgtt 360  
 taattttggc aaagaaaagt ttgaagtcaa aaaaggtgat cgaattgcac agctcatttg 420  
 cgaacggatt ttttatccag aaatagaaga agttcaagcc ttggatgaca ccgaaagggg 480  
 ttcaggaggt tttgggtcca ctggaaagaa ttaaaattta tgccaagaac agaaaacaag 540  
 aagtcatacc tttttcttaa aaaaaaaaaa aaagtttttg cttcaagtgt tttgggtgtt 600  
 tgcacttctg taaacttact agctttacct tctaaaagta ctgcattttt tacttttttt 660  
 tatgatcaag gaaaagatcg ttaaaaaaaaa acacaaagaa gtttttcttt gtgtttggat 720  
 caaaaagaaa ctttggtttt ccgcaattga aggttgtagt taaatctgct ttgtggtgac 780  
 ctgatgtaaa cagtgtcttc ttaaaatcaa atgtaaatac attacagatt aaaaaaaaaa 840  
 gcctgtatth aactcatatg atctcccttc agcaacttat tttgctttta ttgctttaaa 900  
 tcttaagcaa tttttttat tcagtaaaca aattctttca caaggtacaa aatcttgcac 960  
 aagctgaact aaaataaaaa tgaaaaggag agattaaagg tattccttgt tcttcccttc 1020  
 tcttcactag tctaaaaact tctttttaat cttaagattc tttgtgatga gggtagagaa 1080  
 aagaatcctc agttttattt tccactatta atctttcttt tgataaatcc tctattgact 1140  
 gggtagaggt atgtttgtga aagacatgta acttggggat ttgttacttt aggtttgttc 1200  
 ccttgaattt catctcatca ggcaaattgt actagtgtga gttacgagtt ttccctcagt 1260  
 gaagtagcaa taggctgtaa tcaagaaaat atgccattta tagagataag ataaatgaaa 1320  
 taatacttca gccaccagg ttttctgtct cacatacata agcagcattt cattgcagat 1380  
 atgggactga ttctgtggct taccttgatt aacatctttt ggaagttttg ctagtgtgct 1440  
 ttcctttctt tactatgttt ctgagattcc tttgtatcag ggttttgggt gtcacttagg 1500  
 ttttgtccat cagattctgt gagacaccag gcactggttt gaggatgtgg gttatacaca 1560  
 tggagtgttt ctggaactat cagcccactt gaccaccag tttgtggaag cacaggcaag 1620  
 agtgttcttt tctggtgatt ctccaggcca ttaataccc tgcaatgtaa ttgtccctct 1680  
 gtggctcaca tttcattagt gagccatgaa atcaactcag tgggacatag ccagcatttt 1740  
 tgcataccag gttgggctat aaaatatctt tgttgtcaat aaatttttaa tgttttctgt 1800  
 ctaaaaaaaa aaaaaa

&lt;210&gt; 411

&lt;211&gt; 1388

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 411

```

Met Ala Pro Gly Cys Lys Thr Glu Leu Arg Ser Val Thr Asn Gly Gln
 1          5          10          15
Ser Asn Gln Pro Ser Asn Glu Gly Asp Ala Ile Lys Val Phe Val Arg
          20          25          30
Ile Arg Pro Pro Ala Glu Arg Ser Gly Ser Ala Asp Gly Glu Gln Asn
          35          40          45
Leu Cys Leu Ser Val Leu Ser Ser Thr Ser Leu Arg Leu His Ser Asn
          50          55          60
Pro Glu Pro Lys Thr Phe Thr Phe Asp His Val Ala Asp Val Asp Thr
65          70          75          80
Thr Gln Glu Ser Val Phe Ala Thr Val Ala Lys Ser Ile Val Glu Ser
          85          90          95
Cys Met Ser Gly Tyr Asn Gly Thr Ile Phe Ala Tyr Gly Gln Thr Gly
          100          105          110
Ser Gly Lys Thr Phe Thr Met Met Gly Pro Ser Glu Ser Asp Asn Phe
          115          120          125
Ser His Asn Leu Arg Gly Val Ile Pro Arg Ser Phe Glu Tyr Leu Phe
          130          135          140
Ser Leu Ile Asp Arg Glu Lys Glu Lys Ala Gly Ala Gly Lys Ser Phe
          145          150          155          160
Leu Cys Lys Cys Ser Phe Ile Glu Ile Tyr Asn Glu Gln Ile Tyr Asp
          165          170          175
Leu Leu Asp Ser Ala Ser Ala Gly Leu Tyr Leu Arg Glu His Ile Lys
          180          185          190
Lys Gly Val Phe Val Val Gly Ala Val Glu Gln Val Val Thr Ser Ala
          195          200          205
Ala Glu Ala Tyr Gln Val Leu Ser Gly Gly Trp Arg Asn Arg Arg Val
          210          215          220
Ala Ser Thr Ser Met Asn Arg Glu Ser Ser Arg Ser His Ala Val Phe
          225          230          235          240
Thr Ile Thr Ile Glu Ser Met Glu Lys Ser Asn Glu Ile Val Asn Ile
          245          250          255
Arg Thr Ser Leu Leu Asn Leu Val Asp Leu Ala Gly Ser Glu Arg Gln
          260          265          270
Lys Asp Thr His Ala Glu Gly Met Arg Leu Lys Glu Ala Gly Asn Ile
          275          280          285
Asn Arg Ser Leu Ser Cys Leu Gly Gln Val Ile Thr Ala Leu Val Asp
          290          295          300
Val Gly Asn Gly Lys Gln Arg His Val Cys Tyr Arg Asp Ser Lys Leu
          305          310          315          320
Thr Phe Leu Leu Arg Asp Ser Leu Gly Gly Asn Ala Lys Thr Ala Ile
          325          330          335
Ile Ala Asn Val His Pro Gly Ser Arg Cys Phe Gly Glu Thr Leu Ser
          340          345          350
Thr Leu Asn Phe Ala Gln Arg Ala Lys Leu Ile Lys Asn Lys Ala Val
          355          360          365
Val Asn Glu Asp Thr Gln Gly Asn Val Ser Gln Leu Gln Ala Glu Val
          370          375          380
Lys Arg Leu Lys Glu Gln Leu Ala Glu Leu Ala Ser Gly Gln Thr Pro
          385          390          395          400
Pro Glu Ser Phe Leu Thr Arg Asp Lys Lys Lys Thr Asn Tyr Met Glu
          405          410          415
Tyr Phe Gln Glu Ala Met Leu Phe Phe Lys Lys Ser Glu Gln Glu Lys
          420          425          430
Lys Ser Leu Ile Glu Lys Val Thr Gln Leu Glu Asp Leu Thr Leu Lys
          435          440          445
Lys Glu Lys Phe Ile Gln Ser Asn Lys Met Ile Val Lys Phe Arg Glu

```

450  
 Asp Gln Ile Ile Arg Leu Glu Lys Leu His Lys Glu Ser Arg Gly Gly  
 465  
 Phe Leu Pro Glu Glu Gln Asp Arg Leu Leu Ser Glu Leu Arg Asn Glu  
 470  
 485  
 490  
 495  
 Ile Gln Thr Leu Arg Glu Gln Ile Glu His His Pro Arg Val Ala Lys  
 500  
 505  
 510  
 Tyr Ala Met Glu Asn His Ser Leu Arg Glu Glu Asn Arg Arg Leu Arg  
 515  
 520  
 525  
 Leu Leu Glu Pro Val Lys Arg Ala Gln Glu Met Asp Ala Gln Thr Ile  
 530  
 535  
 540  
 Ala Lys Leu Glu Lys Ala Phe Ser Glu Ile Ser Gly Met Glu Lys Ser  
 545  
 550  
 555  
 560  
 Asp Lys Asn Gln Gln Gly Phe Ser Pro Lys Ala Gln Lys Glu Pro Cys  
 565  
 570  
 575  
 Leu Phe Ala Asn Thr Glu Lys Leu Lys Ala Gln Leu Leu Gln Ile Gln  
 580  
 585  
 590  
 Thr Glu Leu Asn Asn Ser Lys Gln Glu Tyr Glu Glu Phe Lys Glu Leu  
 595  
 600  
 605  
 Thr Arg Lys Arg Gln Leu Glu Leu Glu Ser Glu Leu Gln Ser Leu Gln  
 610  
 615  
 620  
 Lys Ala Asn Leu Asn Leu Glu Asn Leu Leu Glu Ala Thr Lys Ala Cys  
 625  
 630  
 635  
 Lys Arg Gln Glu Val Ser Gln Leu Asn Lys Ile His Ala Glu Thr Leu  
 645  
 650  
 655  
 Lys Ile Ile Thr Thr Pro Thr Lys Ala Tyr Gln Leu His Ser Arg Pro  
 660  
 665  
 670  
 Val Pro Lys Leu Ser Pro Glu Met Gly Ser Phe Gly Ser Leu Tyr Thr  
 675  
 680  
 685  
 Gln Asn Ser Ser Ile Leu Asp Asn Asp Ile Leu Asn Glu Pro Val Pro  
 690  
 695  
 700  
 Pro Glu Met Asn Glu Gln Ala Phe Glu Ala Ile Ser Glu Glu Leu Arg  
 705  
 710  
 715  
 720  
 Thr Val Gln Glu Gln Met Ser Ala Leu Gln Ala Lys Leu Asp Glu Glu  
 725  
 730  
 735  
 Glu His Lys Asn Leu Lys Leu Gln Gln His Val Asp Lys Leu Glu His  
 740  
 745  
 750  
 His Ser Thr Gln Met Gln Glu Leu Phe Ser Ser Glu Arg Ile Asp Trp  
 755  
 760  
 765  
 Thr Lys Gln Gln Glu Glu Leu Ser Gln Leu Asn Val Leu Glu Lys  
 770  
 775  
 780  
 Gln Leu Gln Glu Thr Gln Thr Lys Asn Asp Phe Leu Lys Ser Glu Val  
 785  
 790  
 795  
 800  
 His Asp Leu Arg Val Val Leu His Ser Ala Asp Lys Glu Leu Ser Ser  
 805  
 810  
 815  
 Val Lys Leu Glu Tyr Ser Ser Phe Lys Thr Asn Gln Glu Lys Glu Phe  
 820  
 825  
 830  
 Asn Lys Leu Ser Glu Arg His Met His Val Gln Leu Gln Leu Asp Asn  
 835  
 840  
 845  
 Leu Arg Leu Glu Asn Glu Lys Leu Leu Glu Ser Lys Ala Cys Leu Gln  
 850  
 855  
 860  
 Asp Ser Tyr Asp Asn Leu Gln Glu Ile Met Lys Phe Glu Ile Asp Gln  
 865  
 870  
 875  
 880  
 Leu Ser Arg Asn Leu Gln Asn Phe Lys Lys Glu Asn Glu Thr Leu Lys  
 885  
 890  
 895  
 Ser Asp Leu Asn Asn Leu Met Glu Leu Leu Glu Ala Glu Lys Glu Arg  
 900  
 905  
 910  
 Asn Asn Lys Leu Ser Leu Gln Phe Glu Glu Asp Lys Glu Asn Ser Ser  
 915  
 920  
 925  
 Lys Glu Ile Leu Lys Val Leu Glu Ala Val Arg Gln Glu Lys Gln Lys  
 930  
 935  
 940

Glu Thr Ala Lys Cys Glu Gln Gln Met Ala Lys Val Gln Lys Leu Glu  
 945 950 955 960  
 Glu Ser Leu Leu Ala Thr Glu Lys Val Ile Ser Ser Leu Glu Lys Ser  
 965 970 975  
 Arg Asp Ser Asp Lys Lys Val Val Ala Asp Leu Met Asn Gln Ile Gln  
 980 985 990  
 Glu Leu Arg Thr Ser Val Cys Glu Lys Thr Glu Thr Ile Asp Thr Leu  
 995 1000 1005  
 Lys Gln Glu Leu Lys Asp Ile Asn Cys Lys Tyr Asn Ser Ala Leu Val  
 1010 1015 1020  
 Asp Arg Glu Glu Ser Arg Val Leu Ile Lys Lys Gln Glu Val Asp Ile  
 1025 1030 1035 1040  
 Leu Asp Leu Lys Glu Thr Leu Arg Leu Arg Ile Leu Ser Glu Asp Ile  
 1045 1050 1055  
 Glu Arg Asp Met Leu Cys Glu Asp Leu Ala His Ala Thr Glu Gln Leu  
 1060 1065 1070  
 Asn Met Leu Thr Glu Ala Ser Lys Lys His Ser Gly Leu Leu Gln Ser  
 1075 1080 1085  
 Ala Gln Glu Glu Leu Thr Lys Lys Glu Ala Leu Ile Gln Glu Leu Gln  
 1090 1095 1100  
 His Lys Leu Asn Gln Lys Lys Glu Glu Val Glu Gln Lys Lys Asn Glu  
 1105 1110 1115 1120  
 Tyr Asn Phe Lys Met Arg Gln Leu Glu His Val Met Asp Ser Ala Ala  
 1125 1130 1135  
 Glu Asp Pro Gln Ser Pro Lys Thr Pro Pro His Phe Gln Thr His Leu  
 1140 1145 1150  
 Ala Lys Leu Leu Glu Thr Gln Glu Glu Ile Glu Asp Gly Arg Ala  
 1155 1160 1165  
 Ser Lys Thr Ser Leu Glu His Leu Val Thr Lys Leu Asn Glu Asp Arg  
 1170 1175 1180  
 Glu Val Lys Asn Ala Glu Ile Leu Arg Met Lys Glu Gln Leu Arg Glu  
 1185 1190 1195 1200  
 Met Glu Asn Leu Arg Leu Glu Ser Gln Gln Leu Ile Glu Lys Asn Trp  
 1205 1210 1215  
 Leu Leu Gln Gly Gln Leu Asp Asp Ile Lys Arg Gln Lys Glu Asn Ser  
 1220 1225 1230  
 Asp Gln Asn His Pro Asp Asn Gln Gln Leu Lys Asn Glu Gln Glu Glu  
 1235 1240 1245  
 Ser Ile Lys Glu Arg Leu Ala Lys Ser Lys Ile Val Glu Glu Met Leu  
 1250 1255 1260  
 Lys Met Lys Ala Asp Leu Glu Glu Val Gln Ser Ala Leu Tyr Asn Lys  
 1265 1270 1275 1280  
 Glu Met Glu Cys Leu Arg Met Thr Asp Glu Val Glu Arg Thr Gln Thr  
 1285 1290 1295  
 Leu Glu Ser Lys Ala Phe Gln Glu Lys Glu Gln Leu Arg Ser Lys Leu  
 1300 1305 1310  
 Glu Glu Met Tyr Glu Glu Arg Glu Arg Thr Ser Gln Glu Met Glu Met  
 1315 1320 1325  
 Leu Arg Lys Gln Val Glu Cys Leu Ala Glu Glu Asn Gly Lys Leu Val  
 1330 1335 1340  
 Gly His Gln Asn Leu His Gln Lys Ile Gln Tyr Val Val Arg Leu Lys  
 1345 1350 1355 1360  
 Lys Glu Asn Val Arg Leu Ala Glu Glu Thr Glu Lys Leu Arg Ala Glu  
 1365 1370 1375  
 Asn Val Phe Leu Lys Glu Lys Lys Arg Ser Glu Ser  
 1380 1385

<210> 412  
 <211> 4775  
 <212> DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 412

cagtcgcgcg cgggtgcagtc gggaggtgga ggcaccggct gcattgtttt cgggatcgag 60  
gggtgagggc gctatggcac ccggctgcaa aactgagtta cgcagcgtga caaatgggtca 120  
gtctaaccac ccaagtaatg aagggtgatgc catcaaagtt tttgtgcgaa ttctgtcctcc 180  
tgcagaaaaga tctgggtcag ctgatggaga gcagaactta tgcttatctg tgctgtcctc 240  
cacgagtctc cggctgcact ccaacctga gcccaagacc ttcacgtttg atcatgttgc 300  
agatgtggat accactcagg aatctgtatt tgcaactgtg gctaaaagca ttgtggagtc 360  
ttgcatgagc ggttataatg gtaccatctt tgcatatgga cagactggct cagggaagac 420  
atttactatg atgggaccat ctgaatctga taatttttct cataacctga gaggagtaat 480  
cccacgaagt tttgaatat tgttttctct aattgatcgt gaaaaagaaa aggctggagc 540  
tggaagagt ttcctttgtg agtggttctt tattgaaatc tacaacgagc agatatatga 600  
tctactggac tctgcatcgg ctggactgta cttaagggag catatcaaga agggagtctt 660  
tgtttgttgt gcggtggagc aggtggtaac ctcagctgct gaagcctatc aggtgctgtc 720  
tgagagatgg aggaatagac gtgtggcatc aacatcaatg aacagagaat cgtctaggtc 780  
tcatgccgtc tttacaatta caatagagtc aatggagaaa agtaatgaga ttgtgaatat 840  
acggacctcc ctactcaacc tgggtgattt agcaggatct gaaaggcaaa aagataacca 900  
tgcagaaggg atgagattga aggaagcagg taacataaat cgatcattga gctgcctggg 960  
ccaagtgatt acagcacttg tcgacgtggg taatggaaaa cagagacatg tttgtctacag 1020  
agactccaaa cttaccttct tactacggga ttcccttgga ggtaatgcca aaacagccat 1080  
aattgcaaat gttcatcctg gatccagggt ttttggggaa accctatcaa cacttaactt 1140  
tgctcaaaga gccaaagctg ttaaaaacaa ggcagtagta aatgaagaca cccaaggaaa 1200  
tgtgagccag ctccaagctg aagtgaagag gctcaaagaa caactggcgg agcttgcttc 1260  
aggacagaca ccaccagaaa gcttctctgac cagagacaaa aagaagacta actatatgga 1320  
gtatttccag gaagcaatgt tattctttta gaaatctgaa caggaaaaga agtctctgat 1380  
agaaaaagtt acccaattag aagacctcac cctcaaaaag gaaaaattta ttcaatctaa 1440  
taaaatgatt gtgaaattcc gagaggatca aataatacgc ttggaaaagc tccacaagga 1500  
atcccgggga ggttttctgc ctgaggagca ggaatcgtt ctctcagaat taaggaatga 1560  
gattcaaaact ctgcgagAAC aaatagagca ccacccaga gttgcaaagt atgctatgga 1620  
aatcatttcc ctcaaggagg agaatagaag actgagatta ttagagcctg tgaaaagagc 1680  
tcaagaaatg gatgcccaga ccattgcaaa actagaaaaa gcttttctctg aaataagtg 1740  
catggagaaa agtgacaaaa atcagcaagg attttcacct aaagctcaga aagagccatg 1800  
tttgtttgca aacactgaga agttaaagc acaactcctg caaattcaga cagagctgaa 1860  
taattcaaag caagaatatg aagaattcaa agaacttact agggaaaagg agctagaatt 1920  
ggaatcagag cttcagtcct tgcaaaaagc tcagctgaat aaaattcatg ctgaaacact 2040  
aacaaaagcc tgcaagcggc aagaagtttc ccaaggccta ccaacttcat tcccgaccag taccaaaatt 2100  
taagattata actacaccaa ccaaggtctt atacactcag aattctagca tattagataa 2160  
aagcctgaa atgggaagct ttggtctctt gatgaatgaa caagcttttg aggccatttc 2220  
tgatatatta aatgagccag ttctctctga gagtgctctt caagccaaac tggatgaaga 2280  
tgaagagctt agaacagtgc ttacagcaga tgattggacc ctggaacatc attctacca 2340  
agagcataaa aacctaaagc aggaacaaat ttttgacaaa aaacagcagg aagagcttct 2400  
aatgcaggag cttttctcat cagaaagaat tctagagact caaactaaaa atgacttttt 2460  
ctcacagttg aatgtccttg aaaagcagct ccttcattct gctgacaagg agctttcttc 2520  
gaaaaagtgag gtacatgacc tgcgagtagt caatcaggag aaagaattca acaactttc 2580  
agtgaatttg gaatatagt cattcaaaac agataatctc aggttagaaa acgaaaagct 2640  
tgaaagacac atgcatgtac agcttcaatt ctatgacaac ttacaagaaa taatgaaatt 2700  
gcttgagagc aaagcctgcc tacaggattc aaacttcaaa aaagaaaatg aaactctgaa 2760  
tgagattgac caactttcaa gaaacctcca aaacttcaaa aaagaaacgca ataacaaatt 2820  
atctgatctg aataatttga tggagcttct tgaggcagaa gaaatcttaa aagttcttga 2880  
atcattacag tttgaagaag ataaagaaaa cagttctaaa ggccaagtgt gagcagcaga tggcaaaagt 2940  
ggctgtacgt caggagaaac agaaagagac tgaaaaagtg atcagttccc tggaaaagtc 3000  
acagaaacta gaagagagct tgcttgctac cctcatgaac cagatccagg agctaagaac 3060  
tagagattct gataagaaag ttgtagctga caccctgaaa caagaactga aggacataaa 3120  
atcgggtctg gagaaaacag aaactataga caccctgaaa agagtgttga tcaagaagca 3180  
ttgcaaatat aactctgctt tggttgacag ccttaggctg agaatacttt ctgaggacat 3240  
ggaagtggat attctggatc tgaaagaaac ccttaggctg gagcagctga acatgctcac 3300  
agagagggat atgctctgtg aggcacctgg tcatgccact caggaagaac tgaccaagaa 3360  
agaggcctca aaaaaacact cggggctgct gcagtctgcc aagaaagagg aagtagaaca 3420  
ggaagccctg attcaggaac ttcagcaca gctaaaccaa catgtgatgg attctgctgc 3480  
gaagaagaat gaatataact tcaaaatgag gcaactagaa

```

tgaggatccc cagagtccta agacaccacc tcactttcaa acacatttgg caaaactcct 3540
ggaaacacaa gaacaagaga tagaagatgg aagagcctct aagacttctt tggaacacct 3600
tgtaacaaag ctaaatgaag acagagaagt caaaaatgct gaaatcctca gaatgaagga 3660
gcagttgcgt gaaatggaaa acctacgcct ggaaagtcag cagttaatag agaaaaactg 3720
gctcctgcaa ggtcagctgg atgatattaa aagacaaaag gaaaaacagt atcagaatca 3780
tccagataat caacagctga agaatgaaca agaagaaagt atcaaagaaa gacttgcaaa 3840
aagtaaaata gttgaagaaa tgctgaaaat gaaagcagac ctagaagaag tccaaagtgc 3900
cctttacaac aaagagatgg aatgccttag aatgactgat gaagtcgaac gaacccaaac 3960
tttggagtct aaagcattcc aggaaaaaga acaactgaga tcaaagctgg aagaaatgta 4020
tgaagaaaga gagagaacat cccaggagat ggaaatgtta aggaagcagg tggagtgtct 4080
tgctgaggaa aatggaaagt tggtaggtca ccaaaatttg catcagaaga ttcagtacgt 4140
agtgcgacta aagaaggaaa atgtcaggct tgctgaggag acagaaaagt tgcgtgccga 4200
aaatgtatct ttaaaagaaa agaaaagaag tgaatcttga ggattccggt cagctacct 4260
ggcatcacct tgtttgaaga tgtttcttct cttttacaag taagacctac tcttgccac 4320
ttaggagagc tgaatttatg gaccttaatt attaaatgtt tataagggtg tggtaaccac 4380
ctcaagtttc tgatgaacat tctgcatcca tataaccct gtgacagtca gcagtctgct 4440
attaagtggc ctacttcaag gctttgaatc aacttaaggg aaaacctttt gtctttgtaa 4500
aaataaaagc ctgtagctaa ggtttacagt ggacattagc cagatcattt tcttcttaga 4560
ttatgccata atctcctttg attcttatgg aagtctaac aatatatggt ggttccaaca 4620
cctgcagtga gtttaatgac tgacttagta gcaggtacaa gaagcaaact tggttaata 4680
gattattttt gtattcttac tttaggtatt ttacttgagc attttccatg actgtaaata 4740
aagccatttt ttaagataaa aaaaaaaaaa aaaaaa 4775

```

<210> 413  
 <211> 1056  
 <212> PRT  
 <213> Homo sapiens

<400> 413

Met	Ala	Ser	Gln	Pro	Asn	Ser	Ser	Ala	Lys	Lys	Lys	Glu	Glu	Lys	Gly
1				5					10					15	
Lys	Asn	Ile	Gln	Val	Val	Val	Arg	Cys	Arg	Pro	Phe	Asn	Leu	Ala	Glu
	20							25					30		
Arg	Lys	Ala	Ser	Ala	His	Ser	Ile	Val	Glu	Cys	Asp	Pro	Val	Arg	Lys
	35						40					45			
Glu	Val	Ser	Val	Arg	Thr	Gly	Gly	Leu	Ala	Asp	Lys	Ser	Ser	Arg	Lys
	50					55					60				
Thr	Tyr	Thr	Phe	Asp	Met	Val	Phe	Gly	Ala	Ser	Thr	Lys	Gln	Ile	Asp
65					70					75				80	
Val	Tyr	Arg	Ser	Val	Val	Cys	Pro	Ile	Leu	Asp	Glu	Val	Ile	Met	Gly
				85					90					95	
Tyr	Asn	Cys	Thr	Ile	Phe	Ala	Tyr	Gly	Gln	Thr	Gly	Thr	Gly	Lys	Thr
			100					105					110		
Phe	Thr	Met	Glu	Gly	Glu	Arg	Ser	Pro	Asn	Glu	Glu	Tyr	Thr	Trp	Glu
		115					120					125			
Glu	Asp	Pro	Leu	Ala	Gly	Ile	Ile	Pro	Arg	Thr	Leu	His	Gln	Ile	Phe
	130				135						140				
Glu	Lys	Leu	Thr	Asp	Asn	Gly	Thr	Glu	Phe	Ser	Val	Lys	Val	Ser	Leu
145					150					155					160
Leu	Glu	Ile	Tyr	Asn	Glu	Glu	Leu	Phe	Asp	Leu	Leu	Asn	Pro	Ser	Ser
				165					170					175	
Asp	Val	Ser	Glu	Arg	Leu	Gln	Met	Phe	Asp	Asp	Pro	Arg	Asn	Lys	Arg
			180					185					190		
Gly	Val	Ile	Ile	Lys	Gly	Leu	Glu	Ile	Thr	Val	His	Asn	Lys	Asp	
	195					200						205			
Glu	Val	Tyr	Gln	Ile	Leu	Glu	Lys	Gly	Ala	Ala	Lys	Arg	Thr	Thr	Ala
	210					215					220				
Ala	Thr	Leu	Met	Asn	Ala	Tyr	Ser	Ser	Arg	Ser	His	Ser	Val	Phe	Ser
225					230					235				240	
Val	Thr	Ile	His	Met	Lys	Glu	Thr	Thr	Ile	Asp	Gly	Glu	Glu	Leu	Val
				245					250					255	

Lys Ile Gly Lys Leu Asn Leu Val Asp Leu Ala Gly Ser Glu Asn Ile  
 260 265 270  
 Gly Arg Ser Gly Ala Val Asp Lys Arg Ala Arg Glu Ala Gly Asn Ile  
 275 280 285  
 Asn Gln Ser Leu Leu Thr Leu Gly Arg Val Ile Thr Ala Leu Val Glu  
 290 295 300  
 Arg Thr Pro His Val Pro Tyr Arg Glu Ser Lys Leu Thr Arg Ile Leu  
 305 310 315 320  
 Gln Asp Ser Leu Gly Gly Arg Thr Arg Thr Ser Ile Ile Ala Thr Ile  
 325 330 335  
 Ser Pro Ala Ser Leu Asn Leu Glu Glu Thr Leu Ser Thr Leu Glu Tyr  
 340 345 350  
 Ala His Arg Ala Lys Asn Ile Leu Asn Lys Pro Glu Val Asn Gln Lys  
 355 360 365  
 Leu Thr Lys Lys Ala Leu Ile Lys Glu Tyr Thr Glu Glu Ile Glu Arg  
 370 375 380  
 Leu Lys Arg Asp Leu Ala Ala Arg Glu Lys Asn Gly Val Tyr Ile  
 385 390 395 400  
 Ser Glu Glu Asn Phe Arg Val Met Ser Gly Lys Leu Thr Val Gln Glu  
 405 410 415  
 Glu Gln Ile Val Glu Leu Ile Glu Lys Ile Gly Ala Val Glu Glu Glu  
 420 425 430  
 Leu Asn Arg Val Thr Glu Leu Phe Met Asp Asn Lys Asn Glu Leu Asp  
 435 440 445  
 Gln Cys Lys Ser Asp Leu Gln Asn Lys Thr Gln Glu Leu Glu Thr Thr  
 450 455 460  
 Gln Lys His Leu Gln Glu Thr Lys Leu Gln Leu Val Lys Glu Glu Tyr  
 465 470 475 480  
 Ile Thr Ser Ala Leu Glu Ser Thr Glu Glu Lys Leu His Asp Ala Ala  
 485 490 495  
 Ser Lys Leu Leu Asn Thr Val Glu Glu Thr Thr Lys Asp Val Ser Gly  
 500 505 510  
 Leu His Ser Lys Leu Asp Arg Lys Lys Ala Val Asp Gln His Asn Ala  
 515 520 525  
 Glu Ala Gln Asp Ile Phe Gly Lys Asn Leu Asn Ser Leu Phe Asn Asn  
 530 535 540  
 Met Glu Glu Leu Ile Lys Asp Gly Ser Ser Lys Gln Lys Ala Met Leu  
 545 550 555 560  
 Glu Val His Lys Thr Leu Phe Gly Asn Leu Ser Ser Ser Val Ser  
 565 570 575  
 Ala Leu Asp Thr Ile Thr Thr Val Ala Leu Gly Ser Leu Thr Ser Ile  
 580 585 590  
 Pro Glu Asn Val Ser Thr His Val Ser Gln Ile Phe Asn Met Ile Leu  
 595 600 605  
 Lys Glu Gln Ser Leu Ala Ala Glu Ser Lys Thr Val Leu Gln Glu Leu  
 610 615 620  
 Ile Asn Val Leu Lys Thr Asp Leu Leu Ser Ser Leu Glu Met Ile Leu  
 625 630 635 640  
 Ser Pro Thr Val Val Ser Ile Leu Lys Ile Asn Ser Gln Leu Lys His  
 645 650 655  
 Ile Phe Lys Thr Ser Leu Thr Val Ala Asp Lys Ile Glu Asp Gln Lys  
 660 665 670  
 Lys Glu Leu Asp Gly Phe Leu Ser Ile Leu Cys Asn Asn Leu His Glu  
 675 680 685  
 Leu Gln Glu Asn Thr Ile Cys Ser Leu Val Glu Ser Gln Lys Gln Cys  
 690 695 700  
 Gly Asn Leu Thr Glu Asp Leu Lys Thr Ile Lys Gln Thr His Ser Gln  
 705 710 715 720  
 Glu Leu Cys Lys Leu Met Asn Leu Trp Thr Glu Arg Phe Cys Ala Leu  
 725 730 735  
 Glu Glu Lys Cys Glu Asn Ile Gln Lys Pro Leu Ser Ser Val Gln Glu



Asn	Ile	Gln	Gln	Lys	Ser	Lys	Asp	Ile	Val	Asn	Lys	Met	Thr	Phe	His
	755						760					765			
Ser	Gln	Lys	Phe	Cys	Ala	Asp	Ser	Asp	Gly	Phe	Ser	Gln	Glu	Leu	Arg
	770					775					780				
Asn	Phe	Asn	Gln	Glu	Gly	Thr	Lys	Leu	Val	Glu	Glu	Ser	Val	Lys	His
	785				790					795				800	
Ser	Asp	Lys	Leu	Asn	Gly	Asn	Leu	Glu	Lys	Ile	Ser	Gln	Glu	Thr	Glu
			805						810					815	
Gln	Arg	Cys	Glu	Ser	Leu	Asn	Thr	Arg	Thr	Val	Tyr	Phe	Ser	Glu	Gln
		820						825					830		
Trp	Val	Ser	Ser	Leu	Asn	Glu	Arg	Glu	Gln	Glu	Leu	His	Asn	Leu	Leu
	835						840					845			
Glu	Val	Val	Ser	Gln	Cys	Cys	Glu	Ala	Ser	Ser	Ser	Asp	Ile	Thr	Glu
	850					855					860				
Lys	Ser	Asp	Gly	Arg	Lys	Ala	Ala	His	Glu	Lys	Gln	His	Asn	Ile	Phe
	865				870				875						880
Leu	Asp	Gln	Met	Thr	Ile	Asp	Glu	Asp	Lys	Leu	Ile	Ala	Gln	Asn	Leu
			885					890						895	
Glu	Leu	Asn	Glu	Thr	Ile	Lys	Ile	Gly	Leu	Thr	Lys	Leu	Asn	Cys	Phe
		900						905						910	
Leu	Glu	Gln	Asp	Leu	Lys	Leu	Asp	Ile	Pro	Thr	Gly	Thr	Thr	Pro	Gln
	915						920					925			
Arg	Lys	Ser	Tyr	Leu	Tyr	Pro	Ser	Thr	Leu	Val	Arg	Thr	Glu	Pro	Arg
	930					935					940				
Glu	His	Leu	Leu	Asp	Gln	Leu	Lys	Arg	Lys	Gln	Pro	Glu	Leu	Leu	Met
	945				950				955						960
Met	Leu	Asn	Cys	Ser	Glu	Asn	Asn	Lys	Glu	Glu	Thr	Ile	Pro	Asp	Val
			965						970					975	
Asp	Val	Glu	Glu	Ala	Val	Leu	Gly	Gln	Tyr	Thr	Glu	Glu	Pro	Leu	Ser
		980					985						990		
Gln	Glu	Pro	Ser	Val	Asp	Ala	Gly	Val	Asp	Cys	Ser	Ser	Ile	Gly	Gly
	995					1000							1005		
Val	Pro	Phe	Phe	Gln	His	Lys	Lys	Ser	His	Gly	Lys	Asp	Lys	Glu	Asn
	1010					1015						1020			
Arg	Gly	Ile	Asn	Thr	Leu	Glu	Arg	Ser	Lys	Val	Glu	Glu	Thr	Thr	Glu
	1025				1030				1035						1040
His	Leu	Val	Thr	Lys	Ser	Arg	Leu	Pro	Leu	Arg	Ala	Gln	Ile	Asn	Leu
			1045						1050					1055	

<210> 414  
 <211> 4908  
 <212> DNA  
 <213> Homo sapiens

<400> 414  
 acctgctgac agtcggtcct ccaggccaag cagcgcccga gaggaccagg gagactccgg 60  
 cccctgtcgg cgccaagccc ctccgcccct cacagcgccc aggtccgagg ccgggccttg 120  
 attttttggc ggggaccgtc atggcgctgc agccaaattc gtctgcgaag aagaaagagg 180  
 agaaggggaa gaacatccag gtggtggtga gatgcagacc atttaatttg gcagagcgga 240  
 aagctagcgc ccattcaata gtagaatgtg atcctgtacg aaaagaagtt agtgtacgaa 300  
 ctggaggatt ggctgacaag agctcaagga aaacatacac ttttgatatg gtgtttggag 360  
 catctactaa acagattgat gtttaccgaa gtgttgtttg tccaattctg gatgaagtta 420  
 ttatgggcta taattgcact atctttgcgt atggccaaac tggcactgga aaaactttta 480  
 caatggaagg tgaaagggtca cctaatagaag agtatacctg ggaagaggat cccttggtg 540  
 gtataattcc acgtaccctt catcaaattt ttgagaaact tactgataat ggtactgaat 600  
 tttcagtcac agtgtctctg ttggagatct ataatagaag gctttttgat cttcttaatc 660  
 catcatctga tgtttctgag agactacaga tgtttgatga tccccgtaac aagagaggag 720  
 tgataattaa aggttttagaa gaaattacag tacacaacaa ggatgaagtc tatcaaattt 780  
 tagaaaaggg ggcagcaaaa aggacaactg cagctactct gatgaatgca tactctagtc 840

```

gttccactc agttttctct gttacaatac atatgaaaga aactacgatt gatggagaag 900
agcttggttaa aatcggaag ttgaacttgg ttgatcttgc aggaagtga aacattggcc 960
gttctggagc tgttgataag agagctcggg aagctggaaa tataaatcaa tccctgttga 1020
ctttgggaag ggtcattact gccctttag aaagaacacc tcatgttctc tatcgagaat 1080
ctaaactaac tagaatcctc caggattctc ttggaggggcg tacaagaaca tctataattg 1140
caacaatttc tcctgcatct ctcaatcttg aggaactctc gagtacattg gaatatgctc 1200
atagagcaaa gaacatattg aataagcctg aagtgaatca gaaactcacc aaaaaagctc 1260
ttattaagga gtatacggag gagatagaac gtttaaaacg agatcttgct gcagcccggt 1320
agaaaaatgg agtgatatatt tctgaagaaa atttttagagt catgagtga aaattaactg 1380
ttcaagaaga gcagattgta gaattgattg aaaaaattgg tgctgttgag gaggagctga 1440
atagggttac agagttgitt atggataata aaaatgaact tgaccagtgt aaatctgacc 1500
tgcaaaataa aacacaagaa cttgaaacca ctcaaaaaa tttgcaagaa actaaattac 1560
aacttggttaa agaagaatat atcacatcag ctgttgaaag tactgaggag aaacttcatg 1620
atgctgccag caagctgctt aacacagttg aagaaaactac aaaagatgta tctggtctcc 1680
attccaaact ggatcgtaag aaggcagttg accaacacaa tgcagaagct caggatattt 1740
ttggcaaaaa cctgaatagt ctgtttaata atatggaaga attaattaag gatggcagct 1800
caaagcaaaa ggccatgcta gaagtacata agaccttatt tggtaatctg ctgtcttcca 1860
gtgtctctgc attagatacc attactacag tagcacttgg atctctcaca tctattccag 1920
aaaatgtgtc tactcatggt tctcagattt ttaatatgat actaaaagaa caatcattag 1980
cagcagaaag taaaactgta ctacaggaat tgattaatgt actcaagact gatcttctaa 2040
gttcaactgga aatgatttta tccccactg tgggtgtctat actgaaaatc aatagtcaac 2100
taaagcatat tttcaagact tcattgacag tggccgataa gatagaagat caaaaaaagg 2160
aactagatgg ctttctcagt atactgtgta acaatctaca tgaactacaa gaaaaatacca 2220
ttgttccctt ggttgagtca caaaagcaat gtggaacact aactgaagac ctgaagacaa 2280
taaagcagac ccattcccag gaactttgca agtaaatgaa tctttggaca gagagattct 2340
gtgcttttga ggaaaagtgt gaaaatatac agaaaccact tagtagtgtc caggaaaata 2400
tacagcagaa atctaaggat atagtcaaca aaatgacttt tcacagtcaa aaattttgtg 2460
ctgattctga tggcttctca caggaactca gaaattttta ccaagaaggt acaaaattgg 2520
ttgaagaatc tgtgaaacac tctgataaac tcaatggcaa cctggaaaaa atatctcaag 2580
agactgaaca gagatgtgaa tctctgaaca caagaacagt ttatttttct gaacagtggg 2640
tatcttccct aaatgaaagg gaacaggaac ttcacaactt attggaggtt gtaagccaat 2700
gttgtgaggc ttcaagttca gacatcactg agaaatcaga tggacgtaag gcagctcatg 2760
agaaacagca taacattttt cttgatcaga tgactattga tgaagataaa ttgatagcac 2820
aaaatctaga acttaatgaa accataaaaa ttggtttgac taagcttaat tgctttctgg 2880
aacaggatct gaaactggat atcccaacag gtacgacacc acagaggaaa agttattttat 2940
acccatcaac actggttaaga actgaaccac gtgaacatct ccttgatcag ctgaaaagga 3000
aacagcctga gctgttaatg atgctaaact gttcagaaaa caacaaagaa gagacaattc 3060
cggatgtgga tgtagaagag gcagttctgg ggcagtatac tgaagaacct ctaagtcaag 3120
agccatctgt agatgctggt gtggatttgg catcaattgg cggggttcca tttttccagc 3180
ataaaaaatc acatggaaaa gacaaagaaa acagaggcat taacacactg gagaggtcta 3240
aagtggaga aactacagag cacttggtta caaagagcag attacctctg cgagcccaga 3300
tcaaccttta attcacttgg gggttggcaa ttttattttt aaagaaaact taaaaataaa 3360
acctgaaacc ccagaacttg agccttgtgt atagatttta aaagaatata tatatcagcc 3420
gggcgcggtg gctcatgcct gtaatcccag cactttggga ggctgaggcg ggtgatttgc 3480
ttgagcccag gagtttgaga ccagcctggc caactgggca aaacctcgct tctgttaaaa 3540
attagccggg cgtggtggca cactcctgta atcccagcta ctggggaggc tgaggcacga 3600
gaatcacttg aaccaggaa gcggggttgc agtgagccaa aggtacacca ctacactcca 3660
gcctgggcaa cagagcaaga ctcggtctca aaaacaaaat ttaaaaaaga tataaggcag 3720
tactgtaaat tcagttgaat tttgatattc accattttt ctgtcatccc tatagttcac 3780
tttgatttaa attgggtttc atttgggatt tgcaatgtaa atacgtattt ctagttttca 3840
tataaagtag ttcttttata acaaatgaaa agtatttttc ttgtatatta ttaagtaatg 3900
aatatataag aactgtactc ttctcagctt gagcttaaca taggtaaata tcaccaacat 3960
ctgtccttag aaaggaccat ctcatgtttt ttttcttgc atgacttgtg tttttcttg 4020
catcctccct agacttccct atttcgcttt ctctcggct cactttctcc ctttttattt 4080
ttcaccaaac cattttaga gtcacaaaac ctatccttcc ttattttcag tagtcagaat 4140
tttatctaga aatcttttaa caccttttta gtggttattt ctaaaatcac tgtcaacaat 4200
aaatctaacc ctagtgtat ccctccttta agtattttaa acttgttgcc ccaaatgtga 4260
aagcatttaa ttcctttaag aggcctaact cattcacct gacagagttc acaaaaagcc 4320
cacttttagag tatacattgc tattatggga gaccaccag acatctgact aatggctctg 4380
tgccacactc caagacctgt gcctttttaga gaagctcaca atgattttag gactgtttga 4440
aacttccaat tatgtctata atttatattc ttttgtttac atgatgaaac ttttgttgt 4500

```

```

tgcttggttg tatataatcac aatgtgtaca tgtatctttt tctcgattca aatcttaacc 4560
cttaggactc tggatattttt gatctggcaa ccatatttct ggaagttgag atgtttcagc 4620
ttgaagaacc aaaacagaag gaatatgtac aaagaataaa ttttctgctc acgatgagtt 4680
tagtgtgtaa agtttagaga catctgactt tgatagctaa attaaaccaa accctattga 4740
agaattgaat atatgctact tcaagaaact aaattgatct cgtagaatta tcttaataaa 4800
ataatggcta taatttctct gcaaaatcag atgtcagcat aagcgatgga taatacctaa 4860
taaactgccc tcaagtaaact catggttaat aaatgtggtt tctacatt 4908

```

&lt;210&gt; 415

&lt;211&gt; 398

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 415

```

Met Ala Leu Leu Arg Arg Pro Thr Val Ser Ser Asp Leu Glu Asn Ile
1      5      10     15
Asp Thr Gly Val Asn Ser Lys Val Lys Ser His Val Thr Ile Arg Arg
20     25     30
Thr Val Leu Glu Glu Ile Gly Asn Arg Val Thr Thr Arg Ala Ala Gln
35     40     45
Val Ala Lys Lys Ala Gln Asn Thr Lys Val Pro Val Gln Pro Thr Lys
50     55     60
Thr Thr Asn Val Asn Lys Gln Leu Lys Pro Thr Ala Ser Val Lys Pro
65     70     75     80
Val Gln Met Glu Lys Leu Ala Pro Lys Gly Pro Ser Pro Thr Pro Glu
85     90     95
Asp Val Ser Met Lys Glu Glu Asn Leu Cys Gln Ala Phe Ser Asp Ala
100    105    110
Leu Leu Cys Lys Ile Glu Asp Ile Asp Asn Glu Asp Trp Glu Asn Pro
115    120    125
Gln Leu Cys Ser Asp Tyr Val Lys Asp Ile Tyr Gln Tyr Leu Arg Gln
130    135    140
Leu Glu Val Leu Gln Ser Ile Asn Pro His Phe Leu Asp Gly Arg Asp
145    150    155    160
Ile Asn Gly Arg Met Arg Ala Ile Leu Val Asp Trp Leu Val Gln Val
165    170    175
His Ser Lys Phe Arg Leu Leu Gln Glu Thr Leu Tyr Met Cys Val Gly
180    185    190
Ile Met Asp Arg Phe Leu Gln Val Gln Pro Val Ser Arg Lys Lys Leu
195    200    205
Gln Leu Val Gly Ile Thr Ala Leu Leu Leu Ala Ser Lys Tyr Glu Glu
210    215    220
Met Phe Ser Pro Asn Ile Glu Asp Phe Val Tyr Ile Thr Asp Asn Ala
225    230    235    240
Tyr Thr Ser Ser Gln Ile Arg Glu Met Glu Thr Leu Ile Leu Lys Glu
245    250    255
Leu Lys Phe Glu Leu Gly Arg Pro Leu Pro Leu His Phe Leu Arg Arg
260    265    270
Ala Ser Lys Ala Gly Glu Val Asp Val Glu Gln His Thr Leu Ala Lys
275    280    285
Tyr Leu Met Glu Leu Thr Leu Ile Asp Tyr Asp Met Val His Tyr His
290    295    300
Pro Ser Lys Val Ala Ala Ala Ala Ser Cys Leu Ser Gln Lys Val Leu
305    310    315    320
Gly Gln Gly Lys Trp Asn Leu Lys Gln Gln Tyr Tyr Thr Gly Tyr Thr
325    330    335
Glu Asn Glu Val Leu Glu Val Met Gln His Met Ala Lys Asn Val Val
340    345    350
Lys Val Asn Glu Asn Leu Thr Lys Phe Ile Ala Ile Lys Asn Lys Tyr
355    360    365
Ala Ser Ser Lys Leu Leu Lys Ile Ser Met Ile Pro Gln Leu Asn Ser

```

370 375 380  
 Lys Ala Val Lys Asp Leu Ala Ser Pro Leu Ile Gly Arg Ser  
 385 390 395

<210> 416  
 <211> 1530  
 <212> DNA  
 <213> Homo sapiens

<400> 416  
 aatcctggaa caaggctaca gcgtcgaaaga tccccagcgc tgcgggctcg gagagcagtc 60  
 ctaacggcgc ctcgtacgct agtgctctcc cttttcagtc cgcgtccctc cctggggccgg 120  
 gctggcactc ttgccttccc cgtccctcat ggcgctgctc cgacgcccga cgggtgtccag 180  
 tgatttggag aatattgaca caggagttaa ttctaaagtt aagagtcagtg tgactattag 240  
 gcgaactgtt ttagaagaaa ttggaaatag agttacaacc agagcagcac aagtagctaa 300  
 gaaagctcag aacaccaaag ttccagttca acccaccaaa acaacaaatg tcaacaaaca 360  
 actgaaacct actgcttctg tcaaaccagt acagatggaa aagttggctc caaaggggtcc 420  
 ttctcccaca cctgaggatg tctccatgaa ggaagagaat ctctgccaaag ctttttctga 480  
 tgccttgctc tgcaaaatcg aggacattga taacgaagat tgggagaacc ctcagctctg 540  
 cagtactac gttaaggata tctatcagta tctcaggcag ctggagggtt tgcagtccat 600  
 aaaccacat ttcttagatg gaagagatat aaatggacgc atgcgtgcca tcctagtggg 660  
 ttggctggtg caagtccact ccaagtttag gcttctgcag gagactctgt acatgtgcgt 720  
 tggcattatg gatcgatttt tacaggttca gccagtttcc cggaagaagc ttcaattagt 780  
 tgggattact gctctgctct tggcttccaa gtatgaggag atgttttctc caaatattga 840  
 agactttgtt tacatcacag acaatgctta taccagttcc caaatccgag aaatggaaac 900  
 tctaattttg aaagaattga aatttgagtt gggctcgacc ttgccactac acttcttaag 960  
 gcgagcatca aaagccgggg aggttgatgt tgaacagcac acttttagcca agtatttgat 1020  
 ggagctgact ctcatcgact atgatatggt gcattatcat ctttctaagg tagcagcagc 1080  
 tgcttctgct ttgtctcaga aggttctagg acaaggaaaa tggaaactaa agcagcagta 1140  
 ttacacagga tacacagaga atgaagtatt ggaagtcagtg cagcacatgg ccaagaatgt 1200  
 ggtgaaagta aatgaaaact taactaaatt catcgccatc aagaataagt atgcaagcag 1260  
 caaactcctg aagatcagca tgatccctca gctgaactca aaagccgtca aagaccttgc 1320  
 ctccccactg ataggaaggt cctaggtctgc cgtgggccct ggggatgtgt gcttcattgt 1380  
 gccctttttc ttattggttt agaactcttg attttgtaca tagtcctctg gtctatctca 1440  
 tgaaacctct tctcagacca gttttctaaa catatattga ggaaaaataa agcgattggt 1500  
 ttttcttaag gtaaaaaaaaa aaaaaaaaaa 1530

<210> 417  
 <211> 543  
 <212> PRT  
 <213> Homo sapiens

<400> 417  
 Met Ser Arg Glu Ser Asp Val Glu Ala Gln Gln Ser His Gly Ser Ser  
 1 5 10 15  
 Ala Cys Ser Gln Pro His Gly Ser Val Thr Gln Ser Gln Gly Ser Ser  
 20 25 30  
 Ser Gln Ser Gln Gly Ile Ser Ser Ser Thr Ser Thr Met Pro Asn  
 35 40 45  
 Ser Ser Gln Ser Ser His Ser Ser Ser Gly Thr Leu Ser Ser Leu Glu  
 50 55 60  
 Thr Val Ser Thr Gln Glu Leu Tyr Ser Ile Pro Glu Asp Gln Glu Pro  
 65 70 75 80  
 Glu Asp Gln Glu Pro Glu Glu Pro Thr Pro Ala Pro Trp Ala Arg Leu  
 85 90 95  
 Trp Ala Leu Gln Asp Gly Phe Ala Asn Leu Glu Cys Val Asn Asp Asn  
 100 105 110  
 Tyr Trp Phe Gly Arg Asp Lys Ser Cys Glu Tyr Cys Phe Asp Glu Pro  
 115 120 125  
 Leu Leu Lys Arg Thr Asp Lys Tyr Arg Thr Tyr Ser Lys Lys His Phe

130						135						140			
Arg	Ile	Phe	Arg	Glu	Val	Gly	Pro	Lys	Asn	Ser	Tyr	Ile	Ala	Tyr	Ile
145						150				155					160
Glu	Asp	His	Ser	Gly	Asn	Gly	Thr	Phe	Val	Asn	Thr	Glu	Leu	Val	Gly
				165					170						175
Lys	Gly	Lys	Arg	Arg	Pro	Leu	Asn	Asn	Ser	Glu	Ile	Ala	Leu	Ser	
			180					185					190		
Leu	Ser	Arg	Asn	Lys	Val	Phe	Val	Phe	Phe	Asp	Leu	Thr	Val	Asp	Asp
		195					200					205			
Gln	Ser	Val	Tyr	Pro	Lys	Ala	Leu	Arg	Asp	Glu	Tyr	Ile	Met	Ser	Lys
	210					215					220				
Thr	Leu	Gly	Ser	Gly	Ala	Cys	Gly	Glu	Val	Lys	Leu	Ala	Phe	Glu	Arg
225					230					235					240
Lys	Thr	Cys	Lys	Lys	Val	Ala	Ile	Lys	Ile	Ser	Lys	Arg	Lys	Phe	
			245					250					255		
Ala	Ile	Gly	Ser	Ala	Arg	Glu	Ala	Asp	Pro	Ala	Leu	Asn	Val	Glu	Thr
			260					265					270		
Glu	Ile	Glu	Ile	Leu	Lys	Lys	Leu	Asn	His	Pro	Cys	Ile	Ile	Lys	Ile
	275						280					285			
Lys	Asn	Phe	Phe	Asp	Ala	Glu	Asp	Tyr	Tyr	Ile	Val	Leu	Glu	Leu	Met
	290					295					300				
Glu	Gly	Gly	Glu	Leu	Phe	Asp	Lys	Val	Val	Gly	Asn	Lys	Arg	Leu	Lys
305					310					315					320
Glu	Ala	Thr	Cys	Lys	Leu	Tyr	Phe	Tyr	Gln	Met	Leu	Leu	Ala	Val	Gln
			325						330					335	
Tyr	Leu	His	Glu	Asn	Gly	Ile	Ile	His	Arg	Asp	Leu	Lys	Pro	Glu	Asn
		340						345					350		
Val	Leu	Leu	Ser	Ser	Gln	Glu	Glu	Asp	Cys	Leu	Ile	Lys	Ile	Thr	Asp
	355						360					365			
Phe	Gly	His	Ser	Lys	Ile	Leu	Gly	Glu	Thr	Ser	Leu	Met	Arg	Thr	Leu
	370					375					380				
Cys	Gly	Thr	Pro	Thr	Tyr	Leu	Ala	Pro	Glu	Val	Leu	Val	Ser	Val	Gly
385					390					395					400
Thr	Ala	Gly	Tyr	Asn	Arg	Ala	Val	Asp	Cys	Trp	Ser	Leu	Gly	Val	Ile
			405						410					415	
Leu	Phe	Ile	Cys	Leu	Ser	Gly	Tyr	Pro	Pro	Phe	Ser	Glu	His	Arg	Thr
		420						425					430		
Gln	Val	Ser	Leu	Lys	Asp	Gln	Ile	Thr	Ser	Gly	Lys	Tyr	Asn	Phe	Ile
	435					440						445			
Pro	Glu	Val	Trp	Ala	Glu	Val	Ser	Glu	Lys	Ala	Leu	Asp	Leu	Val	Lys
	450					455					460				
Lys	Leu	Leu	Val	Val	Asp	Pro	Lys	Ala	Arg	Phe	Thr	Thr	Glu	Glu	Ala
465					470				475						480
Leu	Arg	His	Pro	Trp	Leu	Gln	Asp	Glu	Asp	Met	Lys	Arg	Lys	Phe	Gln
			485					490					495		
Asp	Leu	Leu	Ser	Glu	Glu	Asn	Glu	Ser	Thr	Ala	Leu	Pro	Gln	Val	Leu
		500						505					510		
Ala	Gln	Pro	Ser	Thr	Ser	Arg	Lys	Arg	Pro	Arg	Glu	Gly	Glu	Ala	Glu
	515						520					525			
Gly	Ala	Glu	Thr	Thr	Lys	Arg	Pro	Ala	Val	Cys	Ala	Ala	Val	Leu	
	530					535					540				

<210> 418  
 <211> 2547  
 <212> DNA  
 <213> Homo sapiens

<400> 418  
 cttacaaggt acagtcctct gctcaggggg gccaggaggg tcttataggg atcattcacc 60  
 agggctgaat gcttctctga gaagtccttt tcaagtctgag acctctggct gaagaaatct 120

```

gggtggacaa gacgctgcag ttgctggtac ctgtgctgga gcttcgctgt atcaactotg 180
aagggaacggt tgcagtcctat aaggtggaag tagtctcgag tggggtcagg tgcctgcagc 240
gctcggcact gtgggcagaa gaacctgtcc tcccgcgccg ggcccatgag gccgcgcag 300
ttccaacagc ggggataatt gcttcccgc cgcgacgcag catcgcagct tagcgggtctc 360
cttctgggaa cccctgtcgg ccaaaaacccc cacaccgcga gcaaagcccc ggctctcccc 420
cgccacatct ggccggcggc ctatctagcc gtgggtcactc gtggggaaaa gcaaagagag 480
cgtctaacca gactaatgtt gctgattggc tggggagtcg agggggcggg atcaccgcag 540
gggaacccgg gttctaagtt ccgctctccc ttctaaacta caactcccag gaggcattga 600
ggcggcgccg gacggccaca tctgctgctc ctcatgggtc cggcggcagg ggagggggtt 660
ttgattggct gaggggtggag tttgtatctg caggttttagc gccactctgc tggctgaggc 720
tgcggagagt gtgcggctcc aggtgggctc acgcggctcg gatgtctcgg gagtccgatg 780
ttgaggctca gcagtctcat ggcagcagtg cctgttcaca gcccctggc agcgttacct 840
agtcccaagg ctccctctca cagtcccagg gcatatccag ctccctctacc agcacgatgc 900
caaactccag ccagtcctct cactccagct ctgggacact gagctcctta gagacagtgt 960
ccactcagga actctattct attcctgagg accaagaacc tgaggacca gaacctgagg 1020
agcctacccc tgccccctgg gctcgattat gggcccttca ggatggattt gccaatcttg 1080
aatgtgtgaa tgacaactac tggtttggga gggacaaaag ctgtgaatat tgctttgatg 1140
aaccactgct gaaaagaaca gataaatacc gaacatacag caagaaacac ttctcgattt 1200
tcagggaagt gggctctaaa aactcttaca ttgcatacat agaagatcac agtggcaatg 1260
gaacctttgt aaatacagag cttgtaggga aaggaaaacg ccgtcctttg aataacaatt 1320
ctgaaattgc actgtcacta agcagaaata aagtttttgt cttttttgat ctgactgtag 1380
atgatcagtc agtttatect aaggcattaa gagatgaata catcatgtca aaaactcttg 1440
gaagtgggtg ctgtggagag gtaaagctgg ctttcgagag gaaaacatgt aagaaagtag 1500
ccataaagat catcagcaaa aggaagtttg ctattgggtc agcaagagag gcagaccag 1560
ctctcaatgt tgaaacagaa atagaaattt tgaaaaagct aaatcatcct tgcatacatc 1620
agattaaaaa cttttttgat gcagaagatt attatatgtt tttggaattg atggaagggg 1680
gagagctggt tgacaaagtg gtggggaata aacgcctgaa agaagctacc tgcaagctct 1740
atttttacca gatgctcttg gctgtgcagt accttcatga aaacgggtatt atacaccgtg 1800
acttaaagcc agagaatgtt ttactgtcat ctcaagaaga ggactgtctt ataaagatta 1860
ctgatttttg gactccaag attttgggag agacctctct catgagaacc ttatgtggaa 1920
ccccaccta cttggcgccg gaagttcttg tttctgttg gactgctggg tataaccgtg 1980
ctgtggactg ctggagttta ggagttattc tttttatctg ccttagtggg tatccacctt 2040
tctctgagca taggactcaa gtgtcactga aggatcagat caccagtggg aaatacaact 2100
tcattcctga agtctgggca gaagtctcag agaaagctct ggaccttgct aagaagttgt 2160
tggtagtggg tccaaaggca cgttttacga cagaagaagc cttaagacac ccgtggcttc 2220
aggatgaaga catgaagaga aagtttcaag atcttctgtc tgaggaaaat gaatccacag 2280
ctctacccca ggttctagcc cagccttcta ctagtcgaaa gcggccccgt gaaggggaag 2340
ccgaggggtg cgagaccaca aagcgcccag ctgtgtgtgc tgctgtgttg tgaactccgt 2400
ggtttgaaca cgaaagaaat gtaccttctt tcaactctgc atctttcttt tctttgagtc 2460
tgttttttta tagtttgtat ttttaattat ggaataattg ctttttcaca gtcactgatg 2520
tacaattaaa aacctgatgg aacctggg

```

<210> 419  
 <211> 297  
 <212> PRT  
 <213> Homo sapiens

<400> 419  
 Met Glu Asp Tyr Thr Lys Ile Glu Lys Ile Gly Glu Gly Thr Tyr Gly  
 1 5 10 15  
 Val Val Tyr Lys Gly Arg His Lys Thr Thr Gly Gln Val Val Ala Met  
 20 25 30  
 Lys Lys Ile Arg Leu Glu Ser Glu Glu Gly Val Pro Ser Thr Ala  
 35 40 45  
 Ile Arg Glu Ile Ser Leu Leu Lys Glu Leu Arg His Pro Asn Ile Val  
 50 55 60  
 Ser Leu Gln Asp Val Leu Met Gln Asp Ser Arg Leu Tyr Leu Ile Phe  
 65 70 75 80  
 Glu Phe Leu Ser Met Asp Leu Lys Lys Tyr Leu Asp Ser Ile Pro Pro  
 85 90 95  
 Gly Gln Tyr Met Asp Ser Ser Leu Val Lys Ser Tyr Leu Tyr Gln Ile

100	105	110
Leu Gln Gly Ile Val Phe Cys His Ser Arg Arg Val Leu His Arg Asp		
115	120	125
Leu Lys Pro Gln Asn Leu Leu Ile Asp Asp Lys Gly Thr Ile Lys Leu		
130	135	140
Ala Asp Phe Gly Leu Ala Arg Ala Phe Gly Ile Pro Ile Arg Val Tyr		
145	150	155
Thr His Glu Val Val Thr Leu Trp Tyr Arg Ser Pro Glu Val Leu Leu		
165	170	175
Gly Ser Ala Arg Tyr Ser Thr Pro Val Asp Ile Trp Ser Ile Gly Thr		
180	185	190
Ile Phe Ala Glu Leu Ala Thr Lys Lys Pro Leu Phe His Gly Asp Ser		
195	200	205
Glu Ile Asp Gln Leu Phe Arg Ile Phe Arg Ala Leu Gly Thr Pro Asn		
210	215	220
Asn Glu Val Trp Pro Glu Val Glu Ser Leu Gln Asp Tyr Lys Asn Thr		
225	230	235
Phe Pro Lys Trp Lys Pro Gly Ser Leu Ala Ser His Val Lys Asn Leu		
245	250	255
Asp Glu Asn Gly Leu Asp Leu Leu Ser Lys Met Leu Ile Tyr Asp Pro		
260	265	270
Ala Lys Arg Ile Ser Gly Lys Met Ala Leu Asn His Pro Tyr Phe Asn		
275	280	285
Asp Leu Asp Asn Gln Ile Lys Lys Met		
290	295	

<210> 420  
 <211> 1235  
 <212> DNA  
 <213> Homo sapiens

<400> 420

gggggggggg	ggcacttggc	ttcaaagctg	gctcttgga	attgagcgg	gagcgacgcg	60
gttgtttag	ctgccgctgc	ggccgcgcgc	gaataataag	ccgggatcta	ccatacccat	120
tgactaacta	tggagatta	taccaaaata	gagaaaattg	gagaaggtag	ctatggagtt	180
gtgtataagg	gtagacacaa	aactacaggt	caagtggtag	ccatgaaaaa	aatcagacta	240
gaaagtgaag	aggaaggggt	tcctagtact	gcaattcggg	aaatttctct	attaaaggaa	300
cttcgtcatc	caaatatagt	cagtcttcag	gatgtgctta	tgcaggattc	caggttatat	360
ctcatctttg	agtttctttc	catggatctg	aagaaatact	tggattctat	ccctcctggg	420
cagtacatgg	attcttcaact	tgtaaagagt	tatttatacc	aaatcctaca	ggggattgtg	480
ttttgtcact	ctagaagagt	tcttcacaga	gacttaaaac	ctcaaaatct	cttgattgat	540
gacaaaggaa	caattaaact	ggctgatttt	ggccttgcca	gagcttttgg	aatacctatc	600
agagtatata	cacatgaggt	agtaacactc	tggtacagat	ctccagaagt	attgctgggg	660
tcagctcgtt	actcaactcc	agttgacatt	tgagtagatg	gcaccatatt	tgctgaacta	720
gcaactaaga	aaccactttt	ccatggggat	tcagaaattg	atcaactctt	caggattttc	780
agagcttttg	gcactcccaa	taatgaagtg	tgccagaag	tggaatcttt	acaggactat	840
aagaatacat	ttcccaaagt	gaaaccagga	agcctagcat	cccattgtcaa	aaacttggat	900
gaaaattggc	tggatttgct	ctcgaaaatg	ttaactctatg	atccagccaa	acgaatttct	960
ggcaaaatgg	cactgaatca	tccatatattt	aatgatttgg	acaatcagat	taagaagatg	1020
tagctttctg	acaaaaagtt	tccatatgtt	atgtcaacag	atagttgtgt	ttttattgtt	1080
aactcttgct	tatttttgct	ttatatatat	ttctttgtta	tcaaacttca	gctgtacttc	1140
gtcttcta	ttcaaaaata	taacttaaaa	atgtaaatat	tctatatgaa	tttaaatata	1200
attctgtaaa	tgtgaaaaaa	aaaaaaaaaa	aaaaa			1235

<210> 421  
 <211> 240  
 <212> PRT  
 <213> Homo sapiens

<400> 421

Met Glu Asp Tyr Thr Lys Ile Glu Lys Ile Gly Glu Gly Thr Tyr Gly  
 1 5 10 15  
 Val Val Tyr Lys Gly Arg His Lys Thr Thr Gly Gln Val Val Ala Met  
 20 25 30  
 Lys Lys Ile Arg Leu Glu Ser Glu Glu Glu Gly Val Pro Ser Thr Ala  
 35 40 45  
 Ile Arg Glu Ile Ser Leu Leu Lys Glu Leu Arg His Pro Asn Ile Val  
 50 55 60  
 Ser Leu Gln Asp Val Leu Met Gln Asp Ser Arg Leu Tyr Leu Ile Phe  
 65 70 75 80  
 Glu Phe Leu Ser Met Asp Leu Lys Lys Tyr Leu Asp Ser Ile Pro Pro  
 85 90 95  
 Gly Gln Tyr Met Asp Ser Ser Leu Val Lys Val Val Thr Leu Trp Tyr  
 100 105 110  
 Arg Ser Pro Glu Val Leu Leu Gly Ser Ala Arg Tyr Ser Thr Pro Val  
 115 120 125  
 Asp Ile Trp Ser Ile Gly Thr Ile Phe Ala Glu Leu Ala Thr Lys Lys  
 130 135 140  
 Pro Leu Phe His Gly Asp Ser Glu Ile Asp Gln Leu Phe Arg Ile Phe  
 145 150 155 160  
 Arg Ala Leu Gly Thr Pro Asn Asn Glu Val Trp Pro Glu Val Glu Ser  
 165 170 175  
 Leu Gln Asp Tyr Lys Asn Thr Phe Pro Lys Trp Lys Pro Gly Ser Leu  
 180 185 190  
 Ala Ser His Val Lys Asn Leu Asp Glu Asn Gly Leu Asp Leu Leu Ser  
 195 200 205  
 Lys Met Leu Ile Tyr Asp Pro Ala Lys Arg Ile Ser Gly Lys Met Ala  
 210 215 220  
 Leu Asn His Pro Tyr Phe Asn Asp Leu Asp Asn Gln Ile Lys Lys Met  
 225 230 235 240

<210> 422  
 <211> 948  
 <212> DNA  
 <213> Homo sapiens

<400> 422  
 ccattgacta actatggaag attataccaa aatagagaaa attggagaag gtacctatgg 60  
 agttgtgtat aagggttagac acaaaactac aggtcaagtg gtagccatga aaaaaatcag 120  
 actagaaagt gaagaggaag gggttcctag tactgcaatt cgggaaattt ctctattaaa 180  
 ggaacttcgt catccaaata tagtcagtct tcaggatgtg cttatgcagg attccaggtt 240  
 atatctcatc tttgagtttc tttccatgga tctgaagaaa tacttggatt ctatccctcc 300  
 tggtcagtac atggattctt cacttggttaa ggtagtaaca ctctggtaca gatctccaga 360  
 agtattgctg gggtcagctc gttactcaac tccagttgac atttggagta taggcacat 420  
 atttgcgtgaa ctagcaacta agaaaccact tttccatggg gattcagaaa ttgatcaact 480  
 cttcaggatt ttcagagctt tgggcactcc caataatgaa gtgtggccag aagtggaaatc 540  
 tttacaggac tataagaata catttcccaa atggaaaacca ggaagcctag catcccatgt 600  
 caaaaacttg gatgaaaatg gcttggattt gctctcgaaa atgttaatct atgatccagc 660  
 caaacgaatt tctggcaaaa tggcactgaa tcatccatat tttaatgatt tggacaatca 720  
 gattaagaag atgtagcttt ctgacaaaaa gtttccatat gttatgtcaa cagatagttg 780  
 tgtttttatt gtttaactctt gtctattttt gtcttatata tatttctttg ttatcaaaact 840  
 tcagctgtac ttcgtcttct aatttcaaaa atataactta aaaatgtaaa tattctatat 900  
 gaatttaaat ataattctgt aaatgtgaaa aaaaaaaaaa aaaaaaaa 948

<210> 423  
 <211> 433  
 <212> PRT  
 <213> Homo sapiens

<400> 423



Met Ala Leu Arg Val Thr Arg Asn Ser Lys Ile Asn Ala Glu Asn Lys  
 1 5 10 15  
 Ala Lys Ile Asn Met Ala Gly Ala Lys Arg Val Pro Thr Ala Pro Ala  
 20 25 30  
 Ala Thr Ser Lys Pro Gly Leu Arg Pro Arg Thr Ala Leu Gly Asp Ile  
 35 40 45  
 Gly Asn Lys Val Ser Glu Gln Leu Gln Ala Lys Met Pro Met Lys Lys  
 50 55 60  
 Glu Ala Lys Pro Ser Ala Thr Gly Lys Val Ile Asp Lys Lys Leu Pro  
 65 70 75 80  
 Lys Pro Leu Glu Lys Val Pro Met Leu Val Pro Val Pro Val Ser Glu  
 85 90 95  
 Pro Val Pro Glu Pro Glu Pro Glu Pro Glu Pro Glu Pro Val Lys Glu  
 100 105 110  
 Glu Lys Leu Ser Pro Glu Pro Ile Leu Val Asp Thr Ala Ser Pro Ser  
 115 120 125  
 Pro Met Glu Thr Ser Gly Cys Ala Pro Ala Glu Glu Asp Leu Cys Gln  
 130 135 140  
 Ala Phe Ser Asp Val Ile Leu Ala Val Asn Asp Val Asp Ala Glu Asp  
 145 150 155 160  
 Gly Ala Asp Pro Asn Leu Cys Ser Glu Tyr Val Lys Asp Ile Tyr Ala  
 165 170 175  
 Tyr Leu Arg Gln Leu Glu Glu Glu Gln Ala Val Arg Pro Lys Tyr Leu  
 180 185 190  
 Leu Gly Arg Glu Val Thr Gly Asn Met Arg Ala Ile Leu Ile Asp Trp  
 195 200 205  
 Leu Val Gln Val Gln Met Lys Phe Arg Leu Leu Gln Glu Thr Met Tyr  
 210 215 220  
 Met Thr Val Ser Ile Ile Asp Arg Phe Met Gln Asn Asn Cys Val Pro  
 225 230 235 240  
 Lys Lys Met Leu Gln Leu Val Gly Val Thr Ala Met Phe Ile Ala Ser  
 245 250 255  
 Lys Tyr Glu Glu Met Tyr Pro Pro Glu Ile Gly Asp Phe Ala Phe Val  
 260 265 270  
 Thr Asp Asn Thr Tyr Thr Lys His Gln Ile Arg Gln Met Glu Met Lys  
 275 280 285  
 Ile Leu Arg Ala Leu Asn Phe Gly Leu Gly Arg Pro Leu Pro Leu His  
 290 295 300  
 Phe Leu Arg Arg Ala Ser Lys Ile Gly Glu Val Asp Val Glu Gln His  
 305 310 315 320  
 Thr Leu Ala Lys Tyr Leu Met Glu Leu Thr Met Leu Asp Tyr Asp Met  
 325 330 335  
 Val His Phe Pro Pro Ser Gln Ile Ala Ala Gly Ala Phe Cys Leu Ala  
 340 345 350  
 Leu Lys Ile Leu Asp Asn Gly Glu Trp Thr Pro Thr Leu Gln His Tyr  
 355 360 365  
 Leu Ser Tyr Thr Glu Glu Ser Leu Leu Pro Val Met Gln His Leu Ala  
 370 375 380  
 Lys Asn Val Val Met Val Asn Gln Gly Leu Thr Lys His Met Thr Val  
 385 390 395 400  
 Lys Asn Lys Tyr Ala Thr Ser Lys His Ala Lys Ile Ser Thr Leu Pro  
 405 410 415  
 Gln Leu Asn Ser Ala Leu Val Gln Asp Leu Ala Lys Ala Val Ala Lys  
 420 425 430  
 Val

<210> 424  
 <211> 2101  
 <212> DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 424

```

acgaacaggc caataaggag ggagcagtgc ggggtttaaa tctgaggcta ggctggctct 60
tctcggcgtg ctgcggcgga acggctgttg gtttctgctg ggtgtaggtc cttggctggt 120
cgggcctccg gtgttctgct tctcccgcgt gagctgctgc ctggtgaaga ggaagccatg 180
gcgctccgag tcaccaggaa ctcgaaaatt aatgctgaaa ataaggcgaa gatcaacatg 240
gcaggcgcaa agcgcgttcc tacggccctt gctgcaacct ccaagcccgg actgaggcca 300
agaacagctc ttggggacat tggtaacaaa gtcagtgaac aactgcaggc caaaatgcct 360
atgaagaagg aagcaaaacc ttcagctact ggaaaagtca ttgataaaaa actaccaaaa 420
cctcttgaaa aggtacctat gctggtgcca gtgccagtgt ctgagccagt gccagagcca 480
gaacctgagc cagaacctga gcctgttaaa gaagaaaaaac tttcgccctga gcctattttg 540
gttgatactg cctctccaag cccaatggaa acatctggat gtgcccctgc agaagaagac 600
ctgtgtcagg ctttctctga tgaattctt gcagtaaatg atgtggatgc agaagatgga 660
gctgatccaa acctttgtag tgaatatgtg aaagatatct atgcttatct gagacaactt 720
gaggaagagc aagcagtcag accaaaatac ctactgggtc gggaaagtcac tggaaacatg 780
agagccatcc taattgactg gctagtacag gttcaaatga aattcagggtt gttgcaggag 840
accatgtaca tgactgtctc cattattgat cggttcatgc agaataattg tgtgcccaag 900
aagatgctgc agctggttgg tgtcactgcc atgtttattg caagcaaata tgaagaaatg 960
tacctccag aaattggtga ctttgccttt gtgactgaca acacttatac taagcaccaa 1020
atcagacaga tggaaatgaa gattctaaga gctttaaact ttggtctggg tcggcctcta 1080
cctttgcact tccttcggag agcatctaag attggagagg ttgatgtcga gcaacatact 1140
ttggccaaat acctgatgga actaactatg ttggactatg acatgggtgca ctttccctct 1200
tctcaaattg cagcaggagc tttttgctta gcactgaaaa ttctggataa tggatgaatg 1260
acaccaactc tacaacatta cctgtcatat actgaagaat ctcttcttcc agttatgcag 1320
cacctggcta agaattgtag catggtaaat caaggactta caaagcacat gactgtcaag 1380
aacaagtatg ccacatcgaa gcatgctaag atcagcactc taccacagct gaattctgca 1440
ctagttcaag atttagccaa ggctgtggca aaggtgtaac ttgtaaaactt gagttggagt 1500
actatattta caaataaaat tggcaccatg tgccatctgt acatattact gttgcattta 1560
cttttaataa agcttgtggc cccttttact tttttatagc ttaactaatt tgaatgtggt 1620
tacttcctac tgtagggtag cggaaaagtt gtcttaaaag gtatggtggg gatattttta 1680
aaaactcctt ttggtttacc tggggatcca attgatgtat atgtttatat actgggttct 1740
tgttttatat acctggcttt tactttatta atatgagtta ctgaaggatg tggaggtatt 1800
tgaaaatttt acttccatag gacatactgc atgtaagcca agtcatggag aatctgctgc 1860
atagctctat tttaaagtaa aagtctacca ccgaatccct agtccccctg ttttctgttt 1920
cttcttgatg ttgctgccat aattctaagt tatttacttt taccactatt taagttatca 1980
actttagcta gtatcttcaa actttcactt tgaaaaatga gaattttata ttctaagcca 2040
gttttcattt tggttttgtg ttttggttaa taaaacaata ctcaaataca aaaaaaaaaa 2100
a

```

2101

&lt;210&gt; 425

&lt;211&gt; 665

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 425

```

Met Ala Ala Gly Gly Ser Thr Gln Gln Arg Arg Arg Glu Met Ala Ala
1      5      10      15
Ala Ser Ala Ala Ala Ile Ser Gly Ala Gly Arg Cys Arg Leu Ser Lys
20      25      30
Ile Gly Ala Thr Arg Arg Pro Pro Ala Arg Val Arg Val Ala Val
35      40      45
Arg Leu Arg Pro Phe Val Asp Gly Thr Ala Gly Ala Ser Asp Pro Pro
50      55      60
Cys Val Arg Gly Met Asp Ser Cys Ser Leu Glu Ile Ala Asn Trp Arg
65      70      75      80
Asn His Gln Glu Thr Leu Lys Tyr Gln Phe Asp Ala Phe Tyr Gly Glu
85      90      95
Arg Ser Thr Gln Gln Asp Ile Tyr Ala Gly Ser Val Gln Pro Ile Leu
100     105     110
Arg His Leu Leu Glu Gly Gln Asn Ala Ser Val Leu Ala Tyr Gly Pro

```

115	120	125
Thr Gly Ala Gly Lys Thr His Thr Met Leu Gly Ser Pro Glu Gln Pro		
130	135	140
Gly Val Ile Pro Arg Ala Leu Met Asp Leu Leu Gln Leu Thr Arg Glu		
145	150	155
Glu Gly Ala Glu Gly Arg Pro Trp Ala Leu Ser Val Thr Met Ser Tyr		
	165	170
Leu Glu Ile Tyr Gln Glu Lys Val Leu Asp Leu Leu Asp Pro Ala Ser		
	180	185
Gly Asp Leu Val Ile Arg Glu Asp Cys Arg Gly Asn Ile Leu Ile Pro		
	195	200
Gly Leu Ser Gln Lys Pro Ile Ser Ser Phe Ala Asp Phe Glu Arg His		
	210	215
Phe Leu Pro Ala Ser Arg Asn Arg Thr Val Gly Ala Thr Arg Leu Asn		
225	230	235
Gln Arg Ser Ser Arg Ser His Ala Val Leu Leu Val Lys Val Asp Gln		
	245	250
Arg Glu Arg Leu Ala Pro Phe Arg Gln Arg Glu Gly Lys Leu Tyr Leu		
	260	265
Ile Asp Leu Ala Gly Ser Glu Asp Asn Arg Arg Thr Gly Asn Lys Gly		
	275	280
Leu Arg Leu Lys Glu Ser Gly Ala Ile Asn Thr Ser Leu Phe Val Leu		
	290	295
Gly Lys Val Val Asp Ala Leu Asn Gln Gly Leu Pro Arg Val Pro Tyr		
305	310	315
Arg Asp Ser Lys Leu Thr Arg Leu Leu Gln Asp Ser Leu Gly Gly Ser		
	325	330
Ala His Ser Ile Leu Ile Ala Asn Ile Ala Pro Glu Arg Arg Phe Tyr		
	340	345
Leu Asp Thr Val Ser Ala Leu Asn Phe Ala Ala Arg Ser Lys Glu Val		
	355	360
Ile Asn Arg Pro Phe Thr Asn Glu Ser Leu Gln Pro His Ala Leu Gly		
	370	375
Pro Val Lys Leu Ser Gln Lys Glu Leu Leu Gly Pro Pro Glu Ala Lys		
385	390	395
Arg Ala Arg Gly Pro Glu Glu Glu Glu Ile Gly Ser Pro Glu Pro Met		
	405	410
Ala Ala Pro Ala Ser Ala Ser Gln Lys Leu Ser Pro Leu Gln Lys Leu		
	420	425
Ser Ser Met Asp Pro Ala Met Leu Glu Arg Leu Leu Ser Leu Asp Arg		
	435	440
Leu Leu Ala Ser Gln Gly Ser Gln Gly Ala Pro Leu Leu Ser Thr Pro		
	450	455
Lys Arg Glu Arg Met Val Leu Met Lys Thr Val Glu Glu Lys Asp Leu		
465	470	475
Glu Ile Glu Arg Leu Lys Thr Lys Gln Lys Glu Leu Glu Ala Lys Met		
	485	490
Leu Ala Gln Lys Ala Glu Glu Lys Glu Asn His Cys Pro Thr Met Leu		
	500	505
Arg Pro Leu Ser His Arg Thr Val Thr Gly Ala Lys Pro Leu Lys Lys		
	515	520
Ala Val Val Met Pro Leu Gln Leu Ile Gln Glu Gln Ala Ala Ser Pro		
	530	535
Asn Ala Glu Ile His Ile Leu Lys Asn Lys Gly Arg Lys Arg Lys Leu		
545	550	555
Glu Ser Leu Asp Ala Leu Glu Pro Glu Glu Lys Ala Glu Asp Cys Trp		
	565	570
Glu Leu Gln Ile Ser Pro Glu Leu Leu Ala His Gly Arg Gln Lys Ile		
	580	585
Leu Asp Leu Leu Asn Glu Gly Ser Ala Arg Asp Leu Arg Ser Leu Gln		
	595	600
		605

Arg Ile Gly Pro Lys Lys Ala Gln Leu Ile Val Gly Trp Arg Glu Leu  
 610 615 620  
 His Gly Pro Phe Ser Gln Val Glu Asp Leu Glu Arg Val Glu Gly Ile  
 625 630 635 640  
 Thr Gly Lys Gln Met Glu Ser Phe Leu Lys Ala Asn Ile Leu Gly Leu  
 645 650 655  
 Ala Ala Gly Gln Arg Cys Gly Ala Ser  
 660 665

<210> 426  
 <211> 2097  
 <212> DNA  
 <213> Homo sapiens

<400> 426  
 gggcgggccc aaggaggagg tggaatggcc gcggggcggt cgacgcagca gaggcgacgc 60  
 gagatggcgg cagcttcagc ggcggcgac tcaggagctg gtcgctgtcg gctaagcaag 120  
 attggagcta ctgcgtgtcc acctccagct cgcgtaaggg tggctgtgcg actgcggcca 180  
 tttgtggatg gaacagcggg agcaagtgat ccccccgtgt tgcggggcat ggacagctgc 240  
 tctctagaga ttgctaactg gaggaaccac caggagactc tcaaatacca gtttgatgcc 300  
 ttctatgggg agaggagtac tcagcaggac atctatgcag gttcagtgcg gcccatccta 360  
 aggcacttgc tggaagggca gaatgccagt gtgcttgccat atggaccac aggagctggg 420  
 aagacgcaca caatgctggg cagcccagag caacctgggg tgatcccgcg ggctctcatg 480  
 gacctcctgc agctcacaag ggaggagggt gccgaggggc ggccatgggc cctttctgtc 540  
 accatgtctt acctagagat ctaccaggag aaggatttag acctcctgga cctgcttcg 600  
 ggagacctgg taatccgaga agactgccgg gggaatatcc tgattccggg tctctcccag 660  
 aagcccatca gtagctttgc tgattttgag cggcacttcc tgccagccag tcgaaatcgg 720  
 actgtaggag ccacccggct caaccagcgc tctcccgcg gtcagtgtgt gtcctgggtc 780  
 aaggtggacc agcgggaacg tttggcccca tttcgccagc gagagggaaa actctacctg 840  
 attgacttgg ctgggtcaga ggacaaccgg cgcacaggca acaaggccct tcggctaaaa 900  
 gagagtggag ccatcaaac ctccctgttt gtccctggga aagtgtaga tgcgctgaat 960  
 cagggcctcc ctgctgtacc ttatcgggac agcaagctca ctgcctatt gcaggactct 1020  
 ctgggtggct cagcccacag tatccttatt gccaacattg cccctgagag acgcttctac 1080  
 ctagacacag tctccgact caactttgct gccaggtcca aggagtgat caatcggcct 1140  
 tttaccaatg agagcctgca gcctcatgcc ttgggacctg ttaagctgtc tcagaaagaa 1200  
 ttgcttggtc caccagaggg aaagagagcc cgaggccctg aggaagagga gattgggagc 1260  
 cctgagccca tggcagctcc agcctctgcc tcccagaaac tcagccccct acagaagcta 1320  
 agcagcatgg acccgccat gctggagcgc ctctcagct tggaccgtct gcttgcctcc 1380  
 caggggagcc agggggcccc tctgttgagt accccaaagc gagagcggat ggtgctaata 1440  
 aagacagtag aagagaagga cctagagatt gagaggctta agacgaagca aaaagaactg 1500  
 gaggccaaag tgttgcccca gaaggctgag gaaaaggaga accattgtcc cacaatgctc 1560  
 cggccccctt cacatcgcac agtcacaggg gcaaagcccc tgaaaaaggc tgtggtgatg 1620  
 cccctacagc taattcagga gcaggcagca tccccaaatg ccgagatcca catcctgaag 1680  
 aataaaggcc ggaagagaaa gctggagtcc ctggatgcc tagagcctga ggagaaggct 1740  
 gaggactgct gggagctaca gatcagcccc gagctactgg ctcatgggcg ccaaaaaata 1800  
 ctggatctgc tgaacgaagg ctacagccga gatctccgca gtcttcagcg cattggcccc 1860  
 aagaaggccc agctaatacgt gggctggcgg gagctccacg gccccttcag ccagggtggag 1920  
 gacctggaac gcgtggaggg cataacgggg aaacagatgg agtccttct gaaggcaaac 1980  
 atcctgggtc tcgcccggcg ccagcgtgt ggcgcctct gaccgtcgtc tctcactcc 2040  
 gcttttcaa atttttgtat aaccccgtgt tgtgtaaata cagtttttgc tccggtg 2097

<210> 427  
 <211> 209  
 <212> PRT  
 <213> Homo sapiens

<400> 427  
 Met Asn Pro Ser Met Lys Gln Lys Gln Glu Glu Ile Lys Glu Asn Ile  
 1 5 10 15  
 Lys Asn Ser Ser Val Pro Arg Arg Thr Leu Lys Met Ile Gln Pro Ser



Met Ala His Lys Gln Ile Tyr Tyr Ser Asp Lys Tyr Phe Asp Glu His  
 1 5 10 15  
 Tyr Glu Tyr Arg His Val Met Leu Pro Arg Glu Leu Ser Lys Gln Val  
 20 25 30  
 Pro Lys Thr His Leu Met Ser Glu Glu Glu Trp Arg Arg Leu Gly Val  
 35 40 45  
 Gln Gln Ser Leu Gly Trp Val His Tyr Met Ile His Glu Pro Glu Pro  
 50 55 60  
 His Ile Leu Leu Phe Arg Arg Pro Leu Pro Lys Asp Gln Gln Lys  
 65 70 75

<210> 430  
 <211> 627  
 <212> DNA  
 <213> Homo sapiens

<400> 430  
 agtctccggc gagttgttgc ctgggctgga cgtggttttg tctgctgcgc ccgctcttcg 60  
 cgctctcgtt tcattttctg cagcgcgcga cgaggatggc ccacaagcag atctactact 120  
 cggacaagta ctccgacgaa cactacgagt accggcatgt tatgttacct agagaacttt 180  
 ccaaacaagt acctaaaact catctgatgt ctgaagagga gtggaggaga cttggtgtcc 240  
 aacagagtct aggtctgggtt cattacatga ttcattgagc agaaccacat attcttctct 300  
 ttagacgacc tcttccaaaa gatcaacaaa aatgaagttt atctggggat cgtcaaactt 360  
 ttttcaaatt taatgtatat gtgtatataa ggtagtattc agtgaatact tgagaaatgt 420  
 acaaactctt catccatacc tgtgcatgag ctgtattctt cacagcaaca gagctcagtt 480  
 aaatgcaact gcaagtaggt tactgtaaga tgtttaagat aaaagttctt ccagtcagtt 540  
 tttctcttaa gtgcctgttt gagtttactg aaacagttta cttttgttca ataaagtttg 600  
 tatgttgcat ttaaaaaaaaa aaaaaaa 627

<210> 431  
 <211> 620  
 <212> PRT  
 <213> Homo sapiens

<400> 431  
 Met Arg Arg Ser Glu Val Leu Ala Glu Glu Ser Ile Val Cys Leu Gln  
 1 5 10 15  
 Lys Ala Leu Asn His Leu Arg Glu Ile Trp Glu Leu Ile Gly Ile Pro  
 20 25 30  
 Glu Asp Gln Arg Leu Gln Arg Thr Glu Val Val Lys Lys His Ile Lys  
 35 40 45  
 Glu Leu Leu Asp Met Met Ile Ala Glu Glu Glu Ser Leu Lys Glu Arg  
 50 55 60  
 Leu Ile Lys Ser Ile Ser Val Cys Gln Lys Glu Leu Asn Thr Leu Cys  
 65 70 75 80  
 Ser Glu Leu His Val Glu Pro Phe Gln Glu Glu Gly Glu Thr Thr Ile  
 85 90 95  
 Leu Gln Leu Glu Lys Asp Leu Arg Thr Gln Val Glu Leu Met Arg Lys  
 100 105 110  
 Gln Lys Lys Glu Arg Lys Gln Glu Leu Lys Leu Leu Gln Glu Gln Asp  
 115 120 125  
 Gln Glu Leu Cys Glu Ile Leu Cys Met Pro His Tyr Asp Ile Asp Ser  
 130 135 140  
 Ala Ser Val Pro Ser Leu Glu Glu Leu Asn Gln Phe Arg Gln His Val  
 145 150 155 160  
 Thr Thr Leu Arg Glu Thr Lys Ala Ser Arg Arg Glu Glu Phe Val Ser  
 165 170 175  
 Ile Lys Arg Gln Ile Ile Leu Cys Met Glu Glu Leu Asp His Thr Pro  
 180 185 190  
 Asp Thr Ser Phe Glu Arg Asp Val Val Cys Glu Asp Glu Asp Ala Phe

<400> 432

```

gtttgcggtt ggttgttgc ctcggggccg tgtggagtag gtctggacct ggactcacgg 60
ctgcttggag cgtccgccat gaggagaagt gaggtgctgg cggaggagtc catagtatgt 120
ctgcagaaag ccctaaatca ccttcgggaa atatgggagc taattgggat tccagaggac 180
cagcgggttac aaagaactga ggtggttaaag aagcatatca aggaactcct ggatatgatg 240
attgctgaag aggaaagcct gaaggaaaga ctcatcaaaa gcatatccgt ctgtcagaaa 300
gagctgaaca ctctgtgcag cgagttacat gttgagccat ttcaggaaga aggagagacg 360
accatcttgc aactagaaaa agatttgcgc acccaagtgg aattgatgcy aaaacagaaa 420
aaggagagaa aacaggaact gaagctactt caagagcaag atcaagaact gtgcgaaatt 480
ctttgtatgc cccactatga tattgacagt gcctcagtgc ccagcttaga agagctgaac 540
cagttcaggc aacatgtgac aactttgagg gaaacaaagg cttctaggcg tgaggagttt 600
gtcagtataa agagacagat catactgtgt atggaagaat tagaccacac cccagacaca 660
agctttgaaa gagatgtggt gtgtgaagac gaagatgcct tttgtttgtc tttggagaat 720
attgcaacac tacaaaagtt gctacggcag cttgaaatgc agaaatcaca aaatgaagca 780
gtgtgtgagg ggctgcgtac tcaaatecga gagctctggg acaggttgca aatactgaa 840
gaagaaagag aagctgtggc caccattatg tctgggtcaa aggccaaagt ccggaagcg 900
ctgcaattag aagtggatcg gttggaagaa ctgaaaatgc aaaacatgaa gaaagtgatt 960
gaggcaattc gagtggagct ggttcagtac tgggaccagt gcttttatag ccaggagcag 1020
agacaagctt ttgccccttt ctgtgctgag gactacacag aaagtctgct ccagctccac 1080
gatgctgaga ttgtgcggtt aaaaaactac tatgaagttc acaaggaact ctttgaaggt 1140
gtccagaagt gggaagaaac ctggaggcct ttcttagagt ttgagagaaa agcttcagat 1200
ccaaatcgat ttacaaaccg aggaggaaat cttctaaaag aagaaaaaca acgagccaag 1260
ctccagaaaa tgctgcccac gctggaagaa gatgtgaagg cacgaattga attgtgggaa 1320
caggaacatt caaaggcatt tatggtgaat gggcagaaat tcatggagta tgtggcagaa 1380
caatgggaga tgcacgatt ggagaaagag agagccaagc aggaaagaca actgaagaac 1440
aaaaaacaga cagagacaga gatgctgtat ggcagcgtc ctcgaacacc tagcaagcgg 1500
cgaggactgg ctcccaatcc accgggcaaa gcacgtaagc tgaacactac caccatgtcc 1560
aatgctacgg ccaatagtag cattcgccct atctttggag ggacagtcta cactcccc 1620
gtgtctcgac ttctctcttc tggcagcaag ccagtcgctg cttccacctg ttcagggaag 1680
aaaacacccc gtactggcag gcatggagcc aacaaggaga acctggagct caacggcagc 1740
atcctgagtg gtgggtaccc tggctcggcc cccctccagc gcaacttcag cattaattct 1800
gttgccagca cctattctga gtttgcaag gatccgtccc tctctgacag ttccactggt 1860
gggcttcagc gagaactttc aaaggcttcc aaatctgatg ctacttctgg aatcctcaat 1920
tcaaccaaca tccagtcctg agaagccctg atcagtaaac cagctgtggc ttctgtgccc 1980
tagactggac ctaattatat ggggtgact ttagtttttc ttcagcttag gagtgttga 2040
aaccttggcc aggttccatg accatgggcc taacttaaaag atgtgaatga gtgttacagt 2100
tgaaagccca tcataggttt agtggctcta ggagacttgg ttttgactta tatacatgaa 2160
aagtttatgg caagaagtgc aaattttagc atatggggcc tgacttctct accacataat 2220
tctacttgct gaagcatgat caaagcttgt tttatttcac cactgtagga aaatgattga 2280
ctatgcccac cctgggggtt aattttggca tgtatacctg taactagtaa ttaacatctt 2340
ttttgtttag gcatgttcaa ttaatgctgt agctatcata gctttgctct tactgaagc 2400
cttgccccca ccacacagga cagccttctt cctgaagaga atgtctttgt gtgtccgaag 2460
ttgagatggc ctgcccactt ggcaaagagg tgacaggaag gctgggagca gctttgttaa 2520
attgtgttca gttctgttac acagtgcatt gccctttgtt ggggggtatgc atgtatgaac 2580
acacatgctt gtcggaacgc tttctcggcg tttgtccctt ggctctcatc tccccattc 2640
ctgtgcctac tttgectgag ttcttctacc ccgcagttg ccagccagat tgggagtctg 2700
ttgttccaa tgggttgagc tgtctttgtc gtggagatct ggaactttgc acatgtcact 2760
actggggagg tgttctgtct ctagcttcca cgatgaggcg cctcttttac ctatcctctc 2820
aatcactact cttcttgaag cactattatt tattcttccg ctgtctgcct gcagcagtac 2880
tactgtcaac atagtgtaaa tggttctcaa aagcttacca gtgtggactt ggtgttagcc 2940
acgctgttta cctcatacag tacgtgtcct gtttttaaaa tatacaatta ttcttaaaaa 3000
taaattaaaa togtatactt acatttcaaa aaaaaaaaaa aaaa 3044

```

&lt;210&gt; 433

&lt;211&gt; 313

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 433

```

Met Pro Val Ala Gly Ser Glu Leu Pro Arg Arg Pro Leu Pro Pro Ala
1           5           10          15
Ala Gln Glu Arg Asp Ala Glu Pro Arg Pro Pro His Gly Glu Leu Gln

```



BNSDOCID: <WO\_2004055050A2\_1\_>

```

gagaaaattg atgacttcaa agctgaagac tttcagattg aagggtacaa tccgcatcca 1020
actattaaaa tggaaatggc tgtttagggt gctttcaaa gagcttgaag gatattgtca 1080
gtcttttaggg gttgggctgg atgccgaggt aaaagttctt tttgctctaa aagaaaaagg 1140
aactagggtca aaaatctgtc cgtgacctat cagttattaa tttttaagga tgttgccact 1200
ggcaaatgta actgtgccag ttctttccat aataaaaaggc tttgagttaa ctactgagg 1260
gtatctgaca atgctgaggt tatgaacaaa gtgaggagaa tgaaatgtat gtgctcttag 1320
caaaaacatg tatgtgcatt tcaatcccac gtacttataa agaaggttgg tgaatttcac 1380
aagctatttt tggaatattt ttagaatatt ttaagaattt cacaagctat tccctcaa 1440
ctgaggggagc tgagtaacac catcgatcat gatgtagagt gtggttatga actttatagt 1500
tgttttatat gttgctataa taaagaagtg ttctgc 1536

```

&lt;210&gt; 435

&lt;211&gt; 225

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 435

```

Met Asn Ser Asn Val Glu Asn Leu Pro Pro His Ile Ile Arg Leu Val
  1             5             10             15
Tyr Lys Glu Val Thr Thr Leu Thr Ala Asp Pro Pro Asp Gly Ile Lys
             20             25             30
Val Phe Pro Asn Glu Glu Asp Leu Thr Asp Leu Gln Val Thr Ile Glu
             35             40             45
Gly Pro Glu Gly Thr Pro Tyr Ala Gly Gly Leu Phe Arg Met Lys Leu
             50             55             60
Leu Leu Gly Lys Asp Phe Pro Ala Ser Pro Pro Lys Gly Tyr Phe Leu
             65             70             75             80
Thr Lys Ile Phe His Pro Asn Val Gly Ala Asn Gly Glu Ile Cys Val
             85             90             95
Asn Val Leu Lys Arg Asp Trp Thr Ala Glu Leu Gly Ile Arg His Val
             100            105            110
Leu Leu Thr Ile Lys Cys Leu Leu Ile His Pro Asn Pro Glu Ser Ala
             115            120            125
Leu Asn Glu Glu Ala Gly Arg Leu Leu Leu Glu Asn Tyr Glu Glu Tyr
             130            135            140
Ala Ala Arg Ala Arg Leu Leu Thr Glu Ile His Gly Gly Ala Gly Gly
             145            150            155            160
Pro Ser Gly Arg Ala Glu Ala Gly Arg Ala Leu Ala Ser Gly Thr Glu
             165            170            175
Ala Ser Ser Thr Asp Pro Gly Ala Pro Gly Gly Pro Gly Gly Ala Glu
             180            185            190
Gly Pro Met Ala Lys Lys His Ala Gly Glu Arg Asp Lys Lys Leu Ala
             195            200            205
Ala Lys Lys Lys Thr Asp Lys Lys Arg Ala Leu Arg Ala Leu Arg Arg
             210            215            220
Leu
225

```

&lt;210&gt; 436

&lt;211&gt; 890

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 436

```

ggcggaccga agaacgcagg aagggggccg ggggggacccg ccccgggccg gccgcagcca 60
tgaactccaa cgtggagaac ctacccccgc acatcatccg cctggtgtac aaggaggtga 120
cgacactgac cgcagaccca cccgatggca tcaaggtctt tcccaacgag gaggacctca 180
ccgacctcca ggtcaccatc gagggccctg agggggacccc atatgctgga ggtctgttcc 240
gcatgaaact cctgctgggg aaggacttcc ctgcctcccc acccaagggc tacttcctga 300
ccaagatctt ccacccgaac gtgggcgcca atggcgagat ctgcgtcaac gtgctcaaga 360

```

```

gggactggac ggctgagctg ggcattccgac acgtactgct gaccatcaag tgcctgctga 420
tccaccctaa ccccgagtct gcactcaacg aggaggcggg ccgcctgctc ttggagaact 480
acgaggagta tgcggctcgg gcccgtctgc tcacagagat ccacgggggc gccggcgggc 540
ccagcggcag ggccgaagcc ggtcgggccc tggccagtgg cactgaagct tcctccaccg 600
accctggggc cccagggggc cggggagggg ctgagggtcc catggccaag aagcatgctg 660
gcgagcgcga taagaagctg gcgggccaaga aaaagacgga caagaagcgg gcgctgcggg 720
cgctgcggcg gctgtagtgg gctctcttcc tccttccacc gtgaccccaa cctctcctgt 780
cccctccctc caactctgtc tctaagttat ttaaattatg gctggggtcg gggaggggtac 840
agggggcact gggacctgga tttgtttttc taaataaagt tggaaaagca 890

```

&lt;210&gt; 437

&lt;211&gt; 197

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 437

```

Met Gln Arg Ala Ser Arg Leu Lys Arg Glu Leu His Met Leu Ala Thr
  1          5          10          15
Glu Pro Pro Pro Gly Ile Thr Cys Trp Gln Asp Lys Asp Gln Met Asp
          20          25          30
Asp Leu Arg Ala Gln Ile Leu Gly Ala Asn Thr Pro Tyr Glu Lys
          35          40          45
Gly Val Phe Lys Leu Glu Val Ile Ile Pro Glu Arg Tyr Pro Phe Glu
          50          55          60
Pro Pro Gln Ile Arg Phe Leu Thr Pro Ile Tyr His Pro Asn Ile Asp
          65          70          75          80
Ser Ala Gly Arg Ile Cys Leu Asp Val Leu Lys Leu Pro Pro Lys Gly
          85          90          95
Ala Trp Arg Pro Ser Leu Asn Ile Ala Thr Val Leu Thr Ser Ile Gln
          100          105          110
Leu Leu Met Ser Glu Pro Asn Pro Asp Asp Pro Leu Met Ala Asp Ile
          115          120          125
Ser Ser Glu Phe Lys Tyr Asn Lys Pro Ala Phe Leu Lys Asn Ala Arg
          130          135          140
Gln Trp Thr Glu Lys His Ala Arg Gln Lys Gln Lys Ala Asp Glu Glu
          145          150          155          160
Glu Met Leu Asp Asn Leu Pro Glu Ala Gly Asp Ser Arg Val His Asn
          165          170          175
Ser Thr Gln Lys Arg Lys Ala Ser Gln Leu Val Gly Ile Glu Lys Lys
          180          185          190
Phe His Pro Asp Val
          195

```

&lt;210&gt; 438

&lt;211&gt; 928

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 438

```

gcgcgcagcg ctggtacccc gttggtccgc gcgttgctgc gttgtgaggg gtgtcagctc 60
agtgcattccc aggcagctct tagtgtggag cagtgaactg tgtgtggttc cttctacttg 120
gggatcatgc agagagcttc acgtctgaag agagagctgc acatgttagc cacagagcca 180
ccccaggca tcacatgttg gcaagataaa gaccaaattg atgacctgcg agctcaaata 240
ttaggtggag ccaacacacc ttatgagaaa ggtgttttta agctagaagt tatcattcct 300
gagaggtacc catttgaacc tcctcagatc cgatttctca ctccaattta tcaccaaaac 360
attgattctg ctggaaggat ttgtctggat gttctcaaat tgccaccaa aggtgcttg 420
agaccatccc tcaacatcgc aactgtgttg acctctattc agctgtctcat gtcagaaccc 480
aaccctgatg acccgctcat ggctgacata tcctcagaat taaatataa taagccagcc 540
ttcctcaaga atgccagaca gtggacagag aagcatgcaa gacagaaaca aaaggctgat 600
gaggaagaga tgcttgataa tctaccagag gctggtgact ccagagtaca caactcaaca 660

```

cagaaaagga aggccagtca gctagtaggc atagaaaaga aatttcatcc tgatgttttag 720  
 gggacttgct ctgggttcac ttagttaatg tgttctttgc caagggtgatc taagttgcct 780  
 accttgaatt tttttttaa tatatttgat gacataattt ttgtgtagtt tatttatctt 840  
 gtacatatgt attttgaaat cttttaaac tgaaaaataa atagtcattt aatgttgaaa 900  
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaa 928

<210> 439

<211> 91

<212> PRT

<213> Homo sapiens

<400> 439

Met Ala Ala Asn Ala Thr Thr Asn Pro Ser Gln Leu Leu Pro Leu Glu  
 1 5 10 15  
 Leu Val Asp Lys Cys Ile Gly Ser Arg Ile His Ile Val Met Lys Ser  
 20 25 30  
 Asp Lys Glu Ile Val Gly Thr Leu Gly Phe Asp Asp Phe Val Asn  
 35 40 45  
 Met Val Leu Glu Asp Val Thr Glu Phe Glu Ile Thr Pro Glu Gly Arg  
 50 55 60  
 Arg Ile Thr Lys Leu Asp Gln Ile Leu Leu Asn Gly Asn Asn Ile Thr  
 65 70 75 80  
 Met Leu Val Pro Gly Gly Glu Gly Pro Glu Val  
 85 90

<210> 440

<211> 749

<212> DNA

<213> Homo sapiens

<400> 440

atggcgggcta acgctactac caaccgctcg cagctgctgc cgtagagact tgtggacaaa 60  
 tgtataggat caagaattca catcgatgat aagagtgata aggaattgt tggactctt 120  
 ctaggatttg atgactttgt caatatggta ctggaagatg tcaactgagtt tgaaatcaca 180  
 ccagaaggaa gaaggattac taaattagat cagattttgc taaatggaaa taatataaca 240  
 atgctgggttc ctggaggaga aggacctgaa gtgtgaatga gtttccttga cttacactag 300  
 attttgtttt ggcttataat gacaagaaaa tggattttt tttccactt tctaattgtt 360  
 aaatcccata aagctaagtt tcccgtaaaa gggaaagtgc ttgaagatgt gtaccattt 420  
 ttgtaagtta atcatgatta tcttgaaaa agaagaaaag aacttcttct tttgcagatg 480  
 aaaataaagg tgttttttgt taactgtcat tttgtttatt ctactgcagt agccagtggg 540  
 acaaagttag tagttatttt gccacttact tttctgtcat tatatgctta tttgttttgt 600  
 catttacgtg accatttgat tctcaacaa aagttgttcc aaacaaaatg atgaactttg 660  
 atttgaacag gtgcatttaa acaaccggaa atgatcactt agaaaattca attaaaatgc 720  
 tgttgttttg taaaaaaaaa aaaaaaaaa 749

<210> 441

<211> 642

<212> PRT

<213> Homo sapiens

<400> 441

Met Lys Arg Ser Ser Val Ser Ser Gly Gly Ala Gly Arg Leu Ser Met  
 1 5 10 15  
 Gln Glu Leu Arg Ser Gln Asp Val Asn Lys Gln Gly Leu Tyr Thr Pro  
 20 25 30  
 Gln Thr Lys Glu Lys Pro Thr Phe Gly Lys Leu Ser Ile Asn Lys Pro  
 35 40 45  
 Thr Ser Glu Arg Lys Val Ser Leu Phe Gly Lys Arg Thr Ser Gly His  
 50 55 60  
 Gly Ser Arg Asn Ser Gln Leu Gly Ile Phe Ser Ser Ser Glu Lys Ile

65					70					75					80
Lys	Asp	Pro	Arg	Pro	Leu	Asn	Asp	Lys	Ala	Phe	Ile	Gln	Gln	Cys	Ile
				85					90					95	
Arg	Gln	Leu	Cys	Glu	Phe	Leu	Thr	Glu	Asn	Gly	Tyr	Ala	His	Asn	Val
			100					105					110		
Ser	Met	Lys	Ser	Leu	Gln	Ala	Pro	Ser	Val	Lys	Asp	Phe	Leu	Lys	Ile
		115					120					125			
Phe	Thr	Phe	Leu	Tyr	Gly	Phe	Leu	Cys	Pro	Ser	Tyr	Glu	Leu	Pro	Asp
	130					135					140				
Thr	Lys	Phe	Glu	Glu	Glu	Val	Pro	Arg	Ile	Phe	Lys	Asp	Leu	Gly	Tyr
145					150					155				160	
Pro	Phe	Ala	Leu	Ser	Lys	Ser	Ser	Met	Tyr	Thr	Val	Gly	Ala	Pro	His
				165					170					175	
Thr	Trp	Pro	His	Ile	Val	Ala	Ala	Leu	Val	Trp	Leu	Ile	Asp	Cys	Ile
			180					185					190		
Lys	Ile	His	Thr	Ala	Met	Lys	Glu	Ser	Ser	Pro	Leu	Phe	Asp	Asp	Gly
		195					200					205			
Gln	Pro	Trp	Gly	Glu	Glu	Thr	Glu	Asp	Gly	Ile	Met	His	Asn	Lys	Leu
	210					215				220					
Phe	Leu	Asp	Tyr	Thr	Ile	Lys	Cys	Tyr	Glu	Ser	Phe	Met	Ser	Gly	Ala
225					230					235				240	
Asp	Ser	Phe	Asp	Glu	Met	Asn	Ala	Glu	Leu	Gln	Ser	Lys	Leu	Lys	Asp
				245				250						255	
Leu	Phe	Asn	Val	Asp	Ala	Phe	Lys	Leu	Glu	Ser	Leu	Glu	Ala	Lys	Asn
			260					265					270		
Arg	Ala	Leu	Asn	Glu	Gln	Ile	Ala	Arg	Leu	Glu	Gln	Glu	Arg	Glu	Lys
		275					280					285			
Glu	Pro	Asn	Arg	Leu	Glu	Ser	Leu	Arg	Lys	Leu	Lys	Ala	Ser	Leu	Gln
	290					295				300					
Gly	Asp	Val	Gln	Lys	Tyr	Gln	Ala	Tyr	Met	Ser	Asn	Leu	Glu	Ser	His
305					310					315				320	
Ser	Ala	Ile	Leu	Asp	Gln	Lys	Leu	Asn	Gly	Leu	Asn	Glu	Glu	Ile	Ala
				325					330					335	
Arg	Val	Glu	Leu	Glu	Cys	Glu	Thr	Ile	Lys	Gln	Glu	Asn	Thr	Arg	Leu
			340					345					350		
Gln	Asn	Ile	Ile	Asp	Asn	Gln	Lys	Tyr	Ser	Val	Ala	Asp	Ile	Glu	Arg
		355					360					365			
Ile	Asn	His	Glu	Arg	Asn	Glu	Leu	Gln	Gln	Thr	Ile	Asn	Lys	Leu	Thr
	370					375				380					
Lys	Asp	Leu	Glu	Ala	Glu	Gln	Gln	Lys	Leu	Trp	Asn	Glu	Glu	Leu	Lys
385					390					395				400	
Tyr	Ala	Arg	Gly	Lys	Glu	Ala	Ile	Glu	Thr	Gln	Leu	Ala	Glu	Tyr	His
				405				410						415	
Lys	Leu	Ala	Arg	Lys	Leu	Lys	Leu	Ile	Pro	Lys	Gly	Ala	Glu	Asn	Ser
			420					425				430			
Lys	Gly	Tyr	Asp	Phe	Glu	Ile	Lys	Phe	Asn	Pro	Glu	Ala	Gly	Ala	Asn
		435					440					445			
Cys	Leu	Val	Lys	Tyr	Arg	Ala	Gln	Val	Tyr	Val	Pro	Leu	Lys	Glu	Leu
	450					455				460					
Leu	Asn	Glu	Thr	Glu	Glu	Glu	Ile	Asn	Lys	Ala	Leu	Asn	Lys	Lys	Met
465					470					475				480	
Gly	Leu	Glu	Asp	Thr	Leu	Glu	Gln	Leu	Asn	Ala	Met	Ile	Thr	Glu	Ser
				485				490						495	
Lys	Arg	Ser	Val	Arg	Thr	Leu	Lys	Glu	Glu	Val	Gln	Lys	Leu	Asp	Asp
			500					505					510		
Leu	Tyr	Gln	Gln	Lys	Ile	Lys	Glu	Ala	Glu	Glu	Glu	Asp	Glu	Lys	Cys
		515					520					525			
Ala	Ser	Glu	Leu	Glu	Ser	Leu	Glu	Lys	His	Lys	His	Leu	Leu	Glu	Ser
	530					535					540				
Thr	Val	Asn	Gln	Gly	Leu	Ser	Glu	Ala	Met	Asn	Glu	Leu	Asp	Ala	Val
545					550					555				560	

Gln Arg Glu Tyr Gln Leu Val Val Gln Thr Thr Thr Glu Glu Arg Arg  
 565 570 575  
 Lys Val Gly Asn Asn Leu Gln Arg Leu Leu Glu Met Val Ala Thr His  
 580 585 590  
 Val Gly Ser Val Glu Lys His Leu Glu Glu Gln Ile Ala Lys Val Asp  
 595 600 605  
 Arg Glu Tyr Glu Glu Cys Met Ser Glu Asp Leu Ser Glu Asn Ile Lys  
 610 615 620  
 Glu Ile Arg Asp Lys Tyr Glu Lys Lys Ala Thr Leu Ile Lys Ser Ser  
 625 630 635 640  
 Glu Glu

&lt;210&gt; 442

&lt;211&gt; 2150

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 442

```

ctcgagccac gaaggccccc ctgtcctgtc tagcagatac ttgcaagggt tacagaaatt 60
cggctccctgg gtcgtgtcag gaaactggaa aaaagggtcat aagcatgaag cgcagttcag 120
tttcagcggg tggtgctggc cgcctctcca tgcaggaggt aagatcccag gatgtaaata 180
aacaaggcct ctataccctt caaaccaaag agaaaccaac ctttgaaaag ttgagtataa 240
acaaaccgac atctgaaaga aaagtctcgc tatttggaag aagaactagt ggacatggat 300
cccggaaatag tcaacttggg atattttcca gttctgagaa aatcaaggac ccgagaccac 360
ttaatgacaa agcatttcatt cagcagtgtg ttcgacaact ctgtgagttt cttacagaaa 420
atggttatgc acataatgtg tccatgaaat ctctacaagc tccctctgtt aaagacttcc 480
tgaagatcct cacatttctt tatggcttcc tgtgcccctc atacgaactt cctgacacaa 540
agtttgaaga agaggttcca agaactttta aagaccttgg gtatcctttt gcactatcca 600
aaagctccat gtacacagtg ggggctcctc atacatggcc tcacattgtg gcagccttag 660
tttggtctaat agactgcac aagatacata ctgccatgaa agaaagctca cctttatttg 720
atgatgggca gccttgggga gaagaaactg aagatggaat tatgcataat aagttgtttt 780
tggtactacac cataaaatgc tatgagagtt ttatgagtggt tgccgacagc tttgatgaga 840
tgaatgcaga gctgcagtc aaactgaagg atttatttaa tgtggatgct ttttaagctgg 900
aatcattaga agcaaaaaac agagcatttg atgaacagat tgcaagattg gaacaagaaa 960
gagaaaaaga accgaatcgt ctgagatcgt tgagaaaaact gaaggcttcc ttacaaggag 1020
atgttcaaaa gtatcaggca tacatgagca atttggagtc tcattcagcc attcttgacc 1080
agaaattaaa tgggtctcaat gaggaaattg ctgagtaga actagaatgt gaaacaataa 1140
aacaggagaa cactcgacta cagaatatca ttgacaacca gaagtactca gttgcagaca 1200
ttgagcgaat aaatcatgaa agaaatgaat tgcagcagac tattaataaa ttaaccaagg 1260
acctggaagc tgaacaacag aagttgtgga atgaggagtt aaaaatgcc agaggcaaag 1320
aagcgattga aacacaatta gcagagtatc acaaattggc tagaaaaatta aaacttattc 1380
ctaaagggtg tgagaattcc aaaggttatg actttgaaat taagtttaat cccgaggctg 1440
gtgccaactg ccttgtcaaa tacagggtc aagtttatgt acctcttaag gaactcctga 1500
atgaaactga agaagaaatt aataaagccc taaataaaaa aatgggtttg gaggatactt 1560
tagaacaatt gaatgcaatg ataacagaaa gcaagagaag tgtgagaact ctgaaagaag 1620
aagttcaaaa gctggatgat ctttaccac aaaaaattaa ggaagcagag gaagaggatg 1680
aaaaatgtgc cagtgaagct gagtccttgg agaaacacaa gcacctgcta gaaagtactg 1740
ttaaccaggg gctcagtgaa gctatgaatg aattagatgc tgttcagcgg gaataccaac 1800
tagttgtgca aaccacgact gaagaaagac gaaaagtggg aaataacttg caacgtctgt 1860
tagagatggt tgctacacat gttgggtctg tagagaaaca tcttgaggag cagattgcta 1920
aagttgatag agaatatgaa gaatgcatgt cagaagatct ctcggaaaaa attaaagaga 1980
ttagagataa gtatgagaag aaagctactc taattaagtc ttctgaagaa tgaagataaa 2040
atgttgatca tgtatatata tccatagtga ataaaattgt ctcagtaaaa aaaaaaaaaa 2100
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2150

```

&lt;210&gt; 443

&lt;211&gt; 380

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 443

```

Met Gly Ile Gln Gly Leu Ala Lys Leu Ile Ala Asp Val Ala Pro Ser
 1           5           10           15
Ala Ile Arg Glu Asn Asp Ile Lys Ser Tyr Phe Gly Arg Lys Val Ala
 20           25           30
Ile Asp Ala Ser Met Ser Ile Tyr Gln Phe Leu Ile Ala Val Arg Gln
 35           40           45
Gly Gly Asp Val Leu Gln Asn Glu Glu Gly Glu Thr Thr Ser His Leu
 50           55           60
Met Gly Met Phe Tyr Arg Thr Ile Arg Met Met Glu Asn Gly Ile Lys
 65           70           75           80
Pro Val Tyr Val Phe Asp Gly Lys Pro Pro Gln Leu Lys Ser Gly Glu
 85           90           95
Leu Ala Lys Arg Ser Glu Arg Arg Ala Glu Ala Glu Lys Gln Leu Gln
 100          105          110
Gln Ala Gln Ala Ala Gly Ala Glu Gln Glu Val Glu Lys Phe Thr Lys
 115          120          125
Arg Leu Val Lys Val Thr Lys Gln His Asn Asp Glu Cys Lys His Leu
 130          135          140
Leu Ser Leu Met Gly Ile Pro Tyr Leu Asp Ala Pro Ser Glu Ala Glu
 145          150          155          160
Ala Ser Cys Ala Ala Leu Val Lys Ala Gly Lys Val Tyr Ala Ala Ala
 165          170          175
Thr Glu Asp Met Asp Cys Leu Thr Phe Gly Ser Pro Val Leu Met Arg
 180          185          190
His Leu Thr Ala Ser Glu Ala Lys Lys Leu Pro Ile Gln Glu Phe His
 195          200          205
Leu Ser Arg Ile Leu Gln Glu Leu Gly Leu Asn Gln Glu Gln Phe Val
 210          215          220
Asp Leu Cys Ile Leu Leu Gly Ser Asp Tyr Cys Glu Ser Ile Arg Gly
 225          230          235          240
Ile Gly Pro Lys Arg Ala Val Asp Leu Ile Gln Lys His Lys Ser Ile
 245          250          255
Glu Glu Ile Val Arg Arg Leu Asp Pro Asn Lys Tyr Pro Val Pro Glu
 260          265          270
Asn Trp Leu His Lys Glu Ala His Gln Leu Phe Leu Glu Pro Glu Val
 275          280          285
Leu Asp Pro Glu Ser Val Glu Leu Lys Trp Ser Glu Pro Asn Glu Glu
 290          295          300
Glu Leu Ile Lys Phe Met Cys Gly Glu Lys Gln Phe Ser Glu Glu Arg
 305          310          315          320
Ile Arg Ser Gly Val Lys Arg Leu Ser Lys Ser Arg Gln Gly Ser Thr
 325          330          335
Gln Gly Arg Leu Asp Asp Phe Phe Lys Val Thr Gly Ser Leu Ser Ser
 340          345          350
Ala Lys Arg Lys Glu Pro Glu Pro Lys Gly Ser Thr Lys Lys Lys Ala
 355          360          365
Lys Thr Gly Ala Ala Gly Lys Phe Lys Arg Gly Lys
 370          375          380

```

&lt;210&gt; 444

&lt;211&gt; 2265

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 444

```

agtcctgcga ttctgggtgt agagggagca ggggcctgctg gggacctggt gtgggtggag 60
tggggacaag cggtagagaa gggtacgcca gggtcgctga gagactctgt tctccctgga 120
gggactggtt gccatgagag cagccgtctg aggggacgca gcctgcacta cgcgccccaa 180

```

```

gaggctgtgc gtggcgagca ggtcacgtga cgggagcgcg ggctttggaa ggcggctgaa 240
cgtcaggcca cccgccgcta agctgagaag ggagagcgag cttaggaccg cctgcccggg 300
gcaaccccca accaagcttt agccgccgag gccgcgtgtc ccaaaggcca gtcataccctc 360
ctctgtgttg ccatgggaat tcaaggcctg gccaaactaa ttgctgatgt ggccccagct 420
gccatccggg agaatgacat caagagctac tttggccgta aggtggccat tgatgcctct 480
atgagcattt atcagttcct gattgctgtt cgccagggtg gggatgtgct gcagaatgag 540
gaggggtgaga ccaccagcca cctgatgggc atgttctacc gcaccattcg catgatggag 600
aacggcatca agcccgtgta tgtctttgat ggcaagccgc cacagctcaa gtcaggcgag 660
ctggccaaac gcagtgaagc gcgggctgag gcagagaagc agctgcagca ggctcaggct 720
gctggggccg agcaggagggt ggaaaaattc actaagcggc tgggtgaaggt cactaagcag 780
cacaatgatg agtgcaaaca tctgctgagc ctcatgggca tcccttatct tgatgcaccc 840
agtgaggcag aggccagctg tgctgccctg tgaaaggctg gcaaagtcta tgctgcggct 900
accgaggaca tggactgcct caccttcggc agccctgtgc taatgcgaca cctgactgcc 960
agtgaagcca aaaagctgcc aatccaggaa ttccacctga gccggattct gcaggagctg 1020
ggcctgaacc aggaacaggtt tgtggatctg tgcattctgc taggcagtga ctactgtgag 1080
agtatccggg gtattgggccc caagcgggct gtggacctca tccagaagca caagagcatc 1140
gaggagatcg tgcggcgact tgaccccaac aagtaccctg tgccagaaaa ttggctccac 1200
aaggaggctc accagctctt ctltgaacct gagtgctgg acccagagtc tgtggagctg 1260
aagtggagcg agccaaatga agaagagctg atcaagttca tgtgtggtga aaagcagttc 1320
tctgaggagc gaatccgcag tggggtcaag aggctgagta agagccgcca aggcagcacc 1380
cagggccgcc tggatgattt cttcaagggt accggctcac tctcttcagc taagcgcaag 1440
gagccagaac ccaagggatc cactaagaag aaggcaaaaga ctggggcagc agggaagttt 1500
aaaaggggaa aataaatgtg tttcccatc atacctcctt caccacagaa tatttgccgt 1560
cttgtaacct taagagctac agctagagaa accttcacgg ggtggagaga ggattctaag 1620
gcttttctag cgtgaccctt ttcagttagt ctagtccctt ttttacttga tcttaatggc 1680
aagaaggcca cagaggtaact tttccttttt tagctcagga aaatatgtca ggctcaaacc 1740
acttctcagg cagtttaatg gacactaagt ccattgttac atgaaagtga tagatagcaa 1800
caagttttgg agaagagaga gggagataaa agggggagac aaaagatgta cagaaatgat 1860
ttcctggctg gccaaactgt ggccagtggg aggtgatggg ggacctagac tgtgcttttc 1920
tgtcttgctc agccttgacc caccttgaga gagagccacc aggaaggcgc atcttagcag 1980
atgggaggaa ctgctgagag aagatgggca gaaagctgga gcccctggag ttggctgtgt 2040
ctgtgtttgt gactgatttc ttgctgtgtc ttgggtgggc agaaactcga acttgctatg 2100
taatttgtgt ctagttattc agaggagtaa gatgggtgatg ttcacctggc aatcagctga 2160
gttgagactt tggaataaga cactggtttt catgcgctgt ttttgtttta aagttatgaa 2220
gaaaaaagtc aataaaattc taaaagtaaa aaaaaaaaaa aaaaaa 2265

```

<210> 445  
 <211> 277  
 <212> PRT  
 <213> Homo sapiens

<400> 445

```

Met Glu Ala Ala Glu Thr Glu Ala Glu Ala Ala Leu Glu Val Leu
 1          5          10          15
Ala Glu Val Ala Gly Ile Leu Glu Pro Val Gly Leu Gln Glu Ala
 20          25          30
Glu Leu Pro Ala Lys Ile Leu Val Glu Phe Val Val Asp Ser Gln Lys
 35          40          45
Lys Asp Lys Leu Leu Cys Ser Gln Leu Gln Val Ala Asp Phe Leu Gln
 50          55          60
Asn Ile Leu Ala Gln Glu Asp Thr Ala Lys Gly Leu Asp Pro Leu Ala
 65          70          75          80
Ser Glu Asp Thr Ser Arg Gln Lys Ala Ile Ala Ala Lys Glu Gln Trp
 85          90          95
Lys Glu Leu Lys Ala Thr Tyr Arg Glu His Val Glu Ala Ile Lys Ile
100          105          110
Gly Leu Thr Lys Ala Leu Thr Gln Met Glu Glu Ala Gln Arg Lys Arg
115          120          125
Thr Gln Leu Arg Glu Ala Phe Glu Gln Leu Gln Ala Lys Lys Gln Met
130          135          140
Ala Met Glu Lys Arg Arg Ala Val Gln Asn Gln Trp Gln Leu Gln Gln

```



145                      150                      155                      160  
 Glu Lys His Leu Gln His Leu Ala Glu Val Ser Ala Glu Val Arg Glu  
                                  165                      170                      175  
 Arg Lys Thr Gly Thr Gln Gln Glu Leu Asp Arg Val Phe Gln Lys Leu  
                                  180                      185                      190  
 Gly Asn Leu Lys Gln Gln Ala Glu Gln Glu Arg Asp Lys Leu Gln Arg  
                                  195                      200                      205  
 Tyr Gln Thr Phe Leu Gln Leu Leu Tyr Thr Leu Gln Gly Lys Leu Leu  
                                  210                      215                      220  
 Phe Pro Glu Ala Glu Ala Glu Ala Glu Asn Leu Pro Asp Asp Lys Pro  
 225                                   230                                   235                                   240  
 Gln Gln Pro Thr Arg Pro Gln Glu Gln Ser Thr Gly Asp Thr Met Gly  
                                  245                                   250                                   255  
 Arg Asp Pro Gly Val Ser Phe Lys Ala Val Gly Leu Gln Pro Ala Gly  
                                  260                                   265                                   270  
 Asp Val Asn Leu Pro  
                                  275

&lt;210&gt; 446

&lt;211&gt; 1658

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 446

```

ggcagcagct cggctcctgg aaagatggag gcagcggaga cagagcgga agctgcagcc 60
ctagaggtcc tggctgaggt ggcaggcatc ttggaacctg taggcctgca ggaggaggca 120
gaactgccag ccaagatcct ggttgagttt gtggtggact ctcagaagaa agacaagctg 180
ctctgcagcc agcttcagggt agcggatttc ctgcagaaca tcctggctca ggaggacact 240
gctaagggtc tcgacccctt ggcttctgaa gacacgagcc gacagaaggc aattgcagct 300
aaggaacaat ggaaagagct gaaggccacc tacagggagc acgtagaggc catcaaaatt 360
ggcctcacca aggccctgac tcagatggag gaagcccaga ggaaacggac acaactccgg 420
gaagcctttg agcagctcca ggccaagaaa caaatggcca tggagaaacg cagagcagtc 480
cagaaccagt ggcagctaca acaggagaag catctgcagc atctggcgga ggtttctgca 540
gaggtgaggg agcgtaaagc agggactcag caggagcttg acagggtgtt tcagaaactt 600
ggaaacctga agcagcaggc agaacaggag cgggacaagc tgcagaggta tcagaccttc 660
ctccagcttc tgtataacct gcagggttaag ctgttgttcc ctgaggctga ggctgaggca 720
gagaatcttc cagatgataa accccagcag ccgactcgac cccaggagca gagtacagga 780
gacaccatgg ggagagaccc tgggtgtgtcc ttcaaggctg ttggtctaca acctgctgga 840
gatgtaaatt tgccatgact tcctggagga cagcagcatg gagaaaagatc ctagaaaagg 900
cctctgactt ccctcacctc ccaaccatca ttacaggaaa gactgtgaac tctgagttc 960
agcttgattt ctgactacat cccagcaagc tctggcatct gtggattaaa tccctggat 1020
ctctctcagt tgtgtatttg ttcattcttc tatgctggca ggaacaacta ttaatacaga 1080
tactcagaag ccaataacat gacaggagct gggactggtt tgaacacagg gtgtgcagat 1140
ggggaggggg tactggcctt gggcctccta tgatgcagac atggtgaatt taattcaagg 1200
aggaggagaa tgtttttaggc aggtggttat atgtgggaag ataattttat tcatggatcc 1260
aaatgtttgt tgagtccttt ctttgtgcta aggttcttgc ggtgaaccag aattataaca 1320
gtgagctcat ctgactgttt taggatgtac agcctagtgt taacattctt ggtatctttt 1380
tgtgccttat ctaaaacatt tctcgatcac tggtttcaga tgttcattta ttatattctt 1440
ttcaaagatt cagagattgg cttttgtcat ccactattgt atgttttgtt tcattgacct 1500
ctagtatac cttgatcttt ccacttttct gtttccggat tggagaagat gtaccttttt 1560
tgtcaactct tactttttatc agatgatcaa ctcacgtatt tggatcttta tttgttttct 1620
caataaataa tttaagggtta aaaaaaaaaa aaaaaaaaaa 1658
  
```

&lt;210&gt; 447

&lt;211&gt; 277

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 447

Met Glu Ala Ala Glu Thr Glu Ala Glu Ala Ala Ala Leu Glu Val Leu

1	5	10	15
Ala Glu Val	Ala Gly Ile Leu Glu	Pro Val Gly Leu Gln Glu Glu Ala	
	20	25	30
Glu Leu Pro	Ala Lys Ile Leu Val	Glu Phe Val Val Asp Ser Gln Lys	
	35	40	45
Lys Asp Lys	Leu Leu Cys Ser Gln Leu	Gln Val Ala Asp Phe Leu Gln	
	50	55	60
Asn Ile Leu	Ala Gln Glu Asp Thr Ala	Lys Gly Leu Asp Pro Leu Ala	
	65	70	75
Ser Glu Asp	Thr Ser Arg Gln Lys Ala	Ile Ala Ala Lys Glu Gln Trp	
	85	90	95
Lys Glu Leu	Lys Ala Thr Tyr Arg Glu	His Val Glu Ala Ile Lys Ile	
	100	105	110
Gly Leu Thr	Lys Ala Leu Thr Gln Met	Glu Glu Ala Gln Arg Lys Arg	
	115	120	125
Thr Gln Leu	Arg Glu Ala Phe Glu Gln	Leu Gln Ala Lys Lys Gln Met	
	130	135	140
Ala Met Glu	Lys Arg Arg Ala Val Gln	Asn Gln Trp Gln Leu Gln Gln	
	145	150	155
Glu Lys His	Leu Gln His Leu Ala Glu	Val Ser Ala Glu Val Arg Glu	
	165	170	175
Arg Lys Thr	Gly Thr Gln Gln Glu Leu	Asp Arg Val Phe Gln Lys Leu	
	180	185	190
Gly Asn Leu	Lys Gln Gln Ala Glu Gln	Glu Arg Asp Lys Leu Gln Arg	
	195	200	205
Tyr Gln Thr	Phe Leu Gln Leu Tyr Thr	Leu Gln Gly Lys Leu Leu	
	210	215	220
Phe Pro Glu	Ala Glu Ala Glu Asn Leu	Pro Asp Asp Lys Pro	
	225	230	235
Gln Gln Pro	Thr Arg Pro Gln Glu Gln	Ser Thr Gly Asp Thr Met Gly	
	245	250	255
Arg Asp Pro	Gly Val Ser Phe Lys Ala	Val Gly Leu Gln Pro Ala Gly	
	260	265	270
Asp Val Asn	Leu Pro		
	275		

<210> 448  
 <211> 1851  
 <212> DNA  
 <213> Homo sapiens

<400> 448

ggcagcagct	cggctcctgg	aaagatggag	gcagcggaga	cagagcgga	agctgcagcc	60
ctagaggtcc	tggctgaggt	ggcaggcatc	ttggaacctg	taggcctgca	ggaggaggca	120
gaactgccag	ccaagatcct	ggttgagttt	gtggtggact	ctcagaagaa	agacaagctg	180
ctctgcagcc	agcttcaggt	agcggatttc	ctgcagaaca	tcctggctca	ggaggacact	240
gctaagggtc	tcgaccctt	ggcttctgaa	gacacgagcc	gacagaaggc	aattgcagct	300
aaggaacaat	ggaaagagct	gaaggccacc	tacagggagc	acgtagaggc	catcaaaatt	360
ggcctcacca	aggccttgac	tcagatggag	gaagcccaga	ggaaacggac	acaactccg	420
gaagcctttg	agcagctcca	ggccaagaaa	caaattggcca	tggagaaacg	cagagcagtc	480
cagaaccagt	ggcagctaca	acaggagaag	catctgcagc	atctggcgga	ggtttctgca	540
gaggtgaggg	agcgtaagac	agggactcag	caggagcttg	acagggtgtt	tcagaaactt	600
ggaaacctga	agcagcaggc	agaacaggag	cgggacaagc	tgcagaggta	tcagaccttc	660
ctccagcttc	tgtataccct	gcagggttaag	ctgttggtcc	ctgaggctga	ggctgaggca	720
gagaatcttc	cagatgataa	accccagcag	ccgactcgac	cccaggagca	gagtacagga	780
gacaccatgg	ggagagaccc	tggtgtgtcc	ttcaaggctg	ttggtctaca	acctgctgga	840
gatgtaaatt	tgccatgact	tcctggagga	cagcagcatg	gagaaagatc	ctagaaaagg	900
tcagacccaa	ctcaggcctt	ggtgtccctg	gactgcaagt	gtggaaggag	ggaaagcctg	960
gtttacctct	ctctgcatct	gagctctgct	acccatggag	cagatggatg	gtgggaacag	1020
gaaagagctt	atgttacacc	tcattcccat	gcttagccca	cccagagcta	acccctgtct	1080

```

tcttccccag gcctctgact tccctcacct cccaaccatc attacaggaa agactgtgaa 1140
ctcttgagtt cagcttgatt tctgactaca tcccagcaag ctctggcatc tgtggattaa 1200
aatccctgga tctctctcag ttgtgtatct gtcatcttc atatgctggc aggaacaact 1260
attaatacag atactcagaa gccaataaca tgacaggagc tgggactggg ttgaacacag 1320
ggtgtgcaga tggggagggg gtactggcct tgggcctcct atgatgcaga catgggtgaat 1380
ttaattcaag gaggaggaga atgttttagg cagggtggta tatgtgggaa gataatttta 1440
ttcatggatc caaatgtttg ttgagtcctt ttgtgctaag gttcttgccg tgaaccagaa 1500
ttataacagt gagctcatct gactgtttta ggatgtacag cctagtgtta acattcttgg 1560
tatctttttg tgccttatct aaaacatttc tcgatcactg gtttcagatg ttcattttatt 1620
atattctttt caaagattca gagattggct tttgtcatcc actattgtat gttttgtttc 1680
attgacctct agtgatacct tgatctttcc cactttctgt ttccggattg gagaagatgt 1740
accttttttg tcaactctta cttttatcag atgatcaact cacgtatttg gatctttatt 1800
tgttttctca aataaatatt taaggttaaa aaaaaaaaaa aaaaaaaaaa a 1851

```

<210> 449  
 <211> 211  
 <212> PRT  
 <213> Homo sapiens

```

<400> 449
Met Ala Ala Arg Arg Gly Ala Leu Ile Val Leu Glu Gly Val Asp Arg
 1           5           10           15
Ala Gly Lys Ser Thr Gln Ser Arg Lys Leu Val Glu Ala Leu Ser Arg
          20           25           30
Gly Pro Pro Pro Glu Leu Leu Arg Phe Pro Glu Arg Ser Thr Glu Ile
          35           40           45
Gly Lys Leu Leu Ser Ser Tyr Leu Gln Lys Lys Ser Asp Val Glu Asp
          50           55           60
His Ser Val His Leu Leu Phe Ser Ala Asn Arg Trp Glu Gln Val Pro
 65           70           75           80
Leu Ile Lys Glu Lys Leu Ser Gln Gly Val Thr Leu Val Val Asp Arg
          85           90           95
Tyr Ala Phe Ser Gly Val Ala Phe Thr Gly Ala Lys Glu Asn Phe Ser
          100          105          110
Leu Asp Trp Cys Lys Gln Pro Asp Val Gly Leu Pro Lys Pro Asp Leu
          115          120          125
Val Leu Phe Leu Gln Leu Gln Leu Ala Asp Ala Ala Lys Arg Gly Ala
          130          135          140
Phe Gly His Glu Arg Tyr Glu Asn Gly Ala Phe Gln Glu Arg Ala Leu
          145          150          155          160
Arg Cys Phe His Gln Leu Met Lys Asp Thr Thr Leu Asn Trp Lys Met
          165          170          175
Val Asp Ala Ser Lys Arg Leu Glu Ala Val His Glu Glu Leu Arg Val
          180          185          190
Leu Ser Glu Asp Ala Ile Arg Thr Ala Thr Glu Lys Pro Leu Gly Glu
          195          200          205
Leu Trp Lys
          210

```

<210> 450  
 <211> 1000  
 <212> DNA  
 <213> Homo sapiens

```

<400> 450
aattcgccag cggcgcggtg gacagtcatt ggcggccggc gcggggctct catagtgtctg 60
gagggcggtg accgcgccgg gaagagcacg cagagccgca agctgggtga agcgtgtctg 120
cgcgggccac cgccgaact gctccggttc ccggaaagat caactgaaat cggcaaactt 180
ctgagttcct acttgcaaaa gaaaagtgcac gtggaggatc actcgggtgc cctgcttttt 240
tctgcaaatc gctgggaaca agtgccgtta attaaggaaa agttgagcca gggcgtgacc 300

```

```

ctcgctcgtgg acagatacgc attttctggt gtggccttca ccggtgccaa ggagaatttt 360
tccctagact ggtgtaaaca gccagacgtg ggccttccca aaccgcacct ggtcctgttc 420
ctccagttac agctggcgga tgctgccaa ggggagcgt ttggccatga gcgctatgag 480
aacggggctt tccaggagcg ggcgctccgg tgtttccacc agctcatgaa agacacgact 540
ttgaactgga agatgggtgga tgcttccaaa agactcgaag ctgtccatga ggaactccgc 600
gtgctctctg aggacgccat ccgcactgcc acagagaagc cgctggggga gctatggaag 660
tgaccaagg ctgcccactg gagacgcctc tccctgcagt ccccgagag gtgggagact 720
cgcggaaggc cccgtcccca gcggagtcca gacccacaa cttcaggagc tctttcccg 780
cagcagagat ctgcaggctg cctcttctgc cccggagctg ggggtgactg gggacccccg 840
tggtggggac cttggcagtg tggacatgag cagagcgatg gagcagtctc ctgccctctc 900
ccctgtcctg atggcactct gttgtatttt ctactgaag ttcagtgata actctgagca 960
gtttcattgt gatcactgta aatggtaatc agttggaatt 1000

```

&lt;210&gt; 451

&lt;211&gt; 282

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 451

```

Met Pro Leu Leu Thr Gln Gln Ile Gln Asp Glu Asp Asp Gln Tyr Ser
1      5      10      15
Leu Val Ala Ser Leu Asp Asn Val Arg Asn Leu Ser Thr Ile Leu Lys
20      25      30
Ala Ile His Phe Arg Glu His Ala Thr Cys Phe Ala Thr Lys Asn Gly
35      40      45
Ile Lys Val Thr Val Glu Asn Ala Lys Cys Val Gln Ala Asn Ala Phe
50      55      60
Ile Gln Ala Gly Ile Phe Gln Glu Phe Lys Val Gln Glu Glu Ser Val
65      70      75      80
Thr Phe Arg Ile Asn Leu Thr Val Leu Leu Asp Cys Leu Ser Ile Phe
85      90      95
Gly Ser Ser Pro Met Pro Gly Thr Leu Thr Ala Leu Arg Met Cys Tyr
100      105      110
Gln Gly Tyr Gly Tyr Pro Leu Met Leu Phe Leu Glu Glu Gly Gly Val
115      120      125
Val Thr Val Cys Lys Ile Asn Thr Gln Glu Pro Glu Glu Thr Leu Asp
130      135      140
Phe Asp Phe Cys Ser Thr Asn Val Ile Asn Lys Ile Ile Leu Gln Ser
145      150      155      160
Glu Gly Leu Arg Glu Ala Phe Ser Glu Leu Asp Met Thr Ser Glu Val
165      170      175
Leu Gln Ile Thr Met Ser Pro Asp Lys Pro Tyr Phe Arg Leu Ser Thr
180      185      190
Phe Gly Asn Ala Gly Ser Ser His Leu Asp Tyr Pro Lys Asp Ser Asp
195      200      205
Leu Met Glu Ala Phe His Cys Asn Gln Thr Gln Val Asn Arg Tyr Lys
210      215      220
Ile Ser Leu Leu Lys Pro Ser Thr Lys Ala Leu Val Leu Ser Cys Lys
225      230      235      240
Val Ser Ile Arg Thr Asp Asn Arg Gly Phe Leu Ser Leu Gln Tyr Met
245      250      255
Ile Arg Asn Glu Asp Gly Gln Ile Cys Phe Val Glu Tyr Tyr Cys Cys
260      265      270
Pro Asp Glu Glu Val Pro Glu Ser Glu Ser
275      280

```

&lt;210&gt; 452

&lt;211&gt; 1776

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 452

```

ctctatgggt cggcggtctt aagcgctcag ctccgcggtc gtcctccagg tctgtgccgc 60
ctccctccgg tctcggtggc gcggcacgcg cggctctcta ggcctccttc agctctgtgg 120
tgacgggtggc cgaggtggag ggccggtctg aagagtggcg ggactggctt cacttcctcc 180
gcggttcttc ggagccgcct cgctcctctt cagggacttt gctgagaagg gctctcgggc 240
gtccagaccc caccgcaaag gtgtttggcg atccgcccag aagtgtttgg ccccaggagc 300
atccctcggg gccgaatgcg cagtggacga tgccccttct gacccaacag atccaagacg 360
aggatgatca gtacagcctt gtggccagcc ttgacaacgt taggaatctc tccactatct 420
tgaaagctat tcatctccga gaacatgcca cgtgtttcgc aactaaaaat ggtatcaaag 480
taacagtgga aaatgcaaag tgtgtgcaag caaatgcttt tattcaggct ggaatatttc 540
aggagttaa agttcaggaa gagtctgtta cttttcgaat taatttaact gtccttttag 600
actgtttatc tatttttggg tcaagtccca tgccagggac tttaactgca cttcgaatgt 660
gttaccagg ttatggttac cctttgatgc tgctcctgga agaaggagga gtggtgacag 720
tctgcaaaat caatacacag gaacctgagg agacctgga ctttgatttc tgcagacca 780
atgttattaa taaaattatt ctgcagtcag aggggctccg tgaagcattt tctgaattgg 840
atatgacgag tgaagtccta caaattacca tgtctcctga caagccttat ttcaggttat 900
ctacttttgg aaatgcagga agttcccacc ttgactatcc caaagattct gatttgatgg 960
aagcatttca ttgtaatcag acccaagtca acagatacaa gatttcctta ctgaaaccct 1020
ctacaaaggc attagtccca tcttgtaagg tatctattcg gacagataac agaggcttcc 1080
tttcattaca gtatatgatt agaaatgaag atggacaaat atgttttgtg gaattatttc 1140
gctgccctga tgaagaagtt cctgaatctg agtcttgagt atgacaattc actgatattt 1200
atgtgtacat ttatgataga tgaagttctt attctgagta cagtactctt tgtcatttca 1260
tattggattt tctatagaga agaagcacia tggggaagat aggagcaagg tcatgtaccc 1320
taatagttag tatgttttgg aaatccattt tgtagaggcg atgtaataa atgttttcc 1380
gtagtcatag attattcagg actgtccttt agttctgtct tttgaactca tgggaataat 1440
tgtgagtcag cgtaacattt caagagtcta aaggtggccg ggtgtggtgg ctttaaatccc 1500
agcactttgg gaacccgagg tggcggtatc acctgaggtc aggagttcga aaccagcctg 1560
accaacgtgg agaaacccca tctctactaa aaatacaaaa aattagccgg acgtggtggc 1620
acatgcctgt aatgccagct actcgggagg ctgaggcagg aaaattgctt gaaccgggga 1680
ggcggagggt gtggtgagct gaaatgacgc cattgcactc cagcctaggc aataagagca 1740
aaactctgcc tcaaaaaaaaa aaaaaaaaaa aaaaaa 1776

```

&lt;210&gt; 453

&lt;211&gt; 838

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 453

```

Met Ser Leu Gln Val Leu Asn Asp Lys Asn Val Ser Asn Glu Lys Asn
1          5          10          15
Thr Glu Asn Cys Asp Phe Leu Phe Ser Pro Pro Glu Val Thr Gly Arg
20          25          30
Ser Ser Val Leu Arg Val Ser Gln Lys Glu Asn Val Pro Pro Lys Asn
35          40          45
Leu Ala Lys Ala Met Lys Val Thr Phe Gln Thr Pro Leu Arg Asp Pro
50          55          60
Gln Thr His Arg Ile Leu Ser Pro Ser Met Ala Ser Lys Leu Glu Ala
65          70          75          80
Pro Phe Thr Gln Asp Asp Thr Leu Gly Leu Glu Asn Ser His Pro Val
85          90          95
Trp Thr Gln Lys Glu Asn Gln Gln Leu Ile Lys Glu Val Asp Ala Lys
100         105         110
Thr Thr His Gly Ile Leu Gln Lys Pro Val Glu Ala Asp Thr Asp Leu
115         120         125
Leu Gly Asp Ala Ser Pro Ala Phe Gly Ser Gly Ser Ser Ser Glu Ser
130         135         140
Gly Pro Gly Ala Leu Ala Asp Leu Asp Cys Ser Ser Ser Ser Gln Ser
145         150         155         160
Pro Gly Ser Ser Glu Asn Gln Met Val Ser Pro Gly Lys Val Ser Gly
165         170         175

```

Ser Pro Glu Gln Ala Val Glu Glu Asn Leu Ser Ser Tyr Ser Leu Asp  
 180 185 190  
 Arg Arg Val Thr Pro Ala Ser Glu Thr Leu Glu Asp Pro Cys Arg Thr  
 195 200 205  
 Glu Ser Gln His Lys Ala Glu Thr Pro His Gly Ala Glu Glu Glu Cys  
 210 215 220  
 Lys Ala Glu Thr Pro His Gly Ala Glu Glu Glu Cys Arg His Gly Gly  
 225 230 235 240  
 Val Cys Ala Pro Ala Ala Val Ala Thr Ser Pro Pro Gly Ala Ile Pro  
 245 250 255  
 Lys Glu Ala Cys Gly Gly Ala Pro Leu Gln Gly Leu Pro Gly Glu Ala  
 260 265 270  
 Leu Gly Cys Pro Ala Gly Val Gly Thr Pro Val Pro Ala Asp Gly Thr  
 275 280 285  
 Gln Thr Leu Thr Cys Ala His Thr Ser Ala Pro Glu Ser Thr Ala Pro  
 290 295 300  
 Thr Asn His Leu Val Ala Gly Arg Ala Met Thr Leu Ser Pro Gln Glu  
 305 310 315 320  
 Glu Val Ala Ala Gly Gln Met Ala Ser Ser Arg Ser Gly Pro Val  
 325 330 335  
 Lys Leu Glu Phe Asp Val Ser Asp Gly Ala Thr Ser Lys Arg Ala Pro  
 340 345 350  
 Pro Pro Arg Arg Leu Gly Glu Arg Ser Gly Leu Lys Pro Pro Leu Arg  
 355 360 365  
 Lys Ala Ala Val Arg Gln Gln Lys Ala Pro Gln Glu Val Glu Glu Asp  
 370 375 380  
 Asp Gly Arg Ser Gly Ala Gly Glu Asp Pro Pro Met Pro Ala Ser Arg  
 385 390 395 400  
 Gly Ser Tyr His Leu Asp Trp Asp Lys Met Asp Asp Pro Asn Phe Ile  
 405 410 415  
 Pro Phe Gly Gly Asp Thr Lys Ser Gly Cys Ser Glu Ala Gln Pro Pro  
 420 425 430  
 Glu Ser Pro Glu Thr Arg Leu Gly Gln Pro Ala Ala Glu Gln Leu His  
 435 440 445  
 Ala Gly Pro Ala Thr Glu Glu Pro Gly Pro Cys Leu Ser Gln Gln Leu  
 450 455 460  
 His Ser Ala Ser Ala Glu Asp Thr Pro Val Val Gln Leu Ala Ala Glu  
 465 470 475 480  
 Thr Pro Thr Ala Glu Ser Lys Glu Arg Ala Leu Asn Ser Ala Ser Thr  
 485 490 495  
 Ser Leu Pro Thr Ser Cys Pro Gly Ser Glu Pro Val Pro Thr His Gln  
 500 505 510  
 Gln Gly Gln Pro Ala Leu Glu Leu Lys Glu Glu Ser Phe Arg Asp Pro  
 515 520 525  
 Ala Glu Val Leu Gly Thr Gly Ala Glu Val Asp Tyr Leu Glu Gln Phe  
 530 535 540  
 Gly Thr Ser Ser Phe Lys Glu Ser Ala Leu Arg Lys Gln Ser Leu Tyr  
 545 550 555 560  
 Leu Lys Phe Asp Pro Leu Leu Arg Asp Ser Pro Gly Arg Pro Val Pro  
 565 570 575  
 Val Ala Thr Glu Thr Ser Ser Met His Gly Ala Asn Glu Thr Pro Ser  
 580 585 590  
 Gly Arg Pro Arg Glu Ala Lys Leu Val Glu Phe Asp Phe Leu Gly Ala  
 595 600 605  
 Leu Asp Ile Pro Val Pro Gly Pro Pro Pro Gly Val Pro Ala Pro Gly  
 610 615 620  
 Gly Pro Pro Leu Ser Thr Gly Pro Ile Val Asp Leu Leu Gln Tyr Ser  
 625 630 635 640  
 Gln Lys Asp Leu Asp Ala Val Val Lys Ala Thr Gln Glu Glu Asn Arg  
 645 650 655  
 Glu Leu Arg Ser Arg Cys Glu Glu Leu His Gly Lys Asn Leu Glu Leu

ggcgcgcggtg	gcagccaggc	ttggccccc	gcgtggagca	gacgcggacc	cctccttctt	60
ggcgcgcgcg	gcgcgggctc	agagcccggc	aacggggcggg	cgggcagaat	gagtctgcag	120
gtcttaaacg	acaaaaatgt	cagcaatgaa	aaaaatacag	aaaattgcga	cttctctgtt	180
tcgccaccag	aagttaccgg	aagatcgtct	gttcttctgt	tgtcacagaa	agaaaaatgtg	240
ccaccaaga	acctggccaa	agctatgaag	gtgacttttc	agacacctct	gcgggatacca	300
cagagcacga	ggattctaa	tcctagcatg	gccagcaaac	ttgaggctcc	tttctactcag	360
gatgacaccc	ttggactgga	aaactcacac	ccggtctgga	cacagaaaga	gaaccaacag	420
ctcatcaagg	aagtggatgc	caaaactact	catggaattc	tacagaaacc	agtggaggct	480
gacaccgacc	tcctggggga	tgcaagccca	gcctttggga	gtggcagctc	cagcgagtct	540
ggccccagtg	ccctggctga	cctggactgc	tcaagctctt	cccagagccc	aggaagtctt	600
gagaaccaaa	tggttctctc	aggaaaagt	cttggcagcc	ctgagcaagc	cgtggaggaa	660
aaccttagtt	cgtattcctt	agacagaaga	gtgacacccg	cctctgagac	cttagaagac	720
ccttgacagga	cagagtccca	gcacaaagcg	gagactccgc	acggagccga	ggaagaatgc	780
aaagcggaga	ctccgcacgg	agccgaggag	gaatgccggc	acggtgggg	ctgtgctccc	840
gcagcagtg	ccacttcgcc	tcctggtgca	atccctaagg	aagcctgcgg	aggagacccc	900
ctgcagggtc	tgcctggcga	agccctgggc	tgccttgcgg	gtgtgggcac	ccccgtgcca	960
gcagatggca	ctcagaccct	tacctgtgca	cacacctctg	ctcctgagag	cacagcccca	1020
accaaccacc	tggtggctgg	caggctcatg	acacctgagtc	ctcaggaaga	agtggctgca	1080
ggccaaatgg	ccagctctct	gaggagcggg	cctgtaaaac	tagaatttga	tgtatctgat	1140
ggcgccacca	gcaaaagggc	accccacca	aggagactgg	gagagaggtc	cggcctcaag	1200
cctcccttga	ggaaagcagc	agtgaaggcag	caaaaggccc	cgcaggaggt	ggaggaggac	1260
gacggtagga	gcggagcagg	agaggacccc	cccatgcagg	cttctcgggg	ctctttaccac	1320
ctgcagtggg	acaaaatgga	tgacccaaac	ttcatccctg	tcggaggtga	caccaagtct	1380
ggttgcagtg	agggccagcc	ccagaaaagc	cctgagacca	ggctgggcca	gccagcggtt	1440
gaacagttgc	atgctggggc	tgccacggag	gagccaggtc	cctgtctgag	ccagcagctg	1500
cattcagcct	cagcggagga	cacgcctgtg	gtgcagttgg	cagccgagac	ccaacacgca	1560
gagagcaagg	agagagcctt	gaactctgcc	agcacctcgc	ttcccacaag	ctgtccaggc	1620
agtgagccag	tgcccaccca	tcagcagggg	cagcctgcct	tggagctgaa	agaggagagc	1680
ttcagagacc	ccgctagagt	tttaggcacg	ggcgcggagg	tggattacct	ggagcagttt	1740
ggaacttcct	cgtttaagga	gtcggccttg	aggaagcagt	cctataacct	caaqttcgac	1800

```

ccccctcctga gggacagtc tggtagacca gtgcccgtgg ccaccgagac cagcagcatg 1860
cacggtgcaa atgagactcc ctcaggacgt ccgcgggaag ccaagcttgt ggagttcgat 1920
ttcttgggag cactggacat tcctgtgcca ggcccacccc caggtgttcc cgcgccctggg 1980
ggcccacccc tgtccacggg acctatagtg gacctgctcc agtacagcca gaaggacctg 2040
gatgcagtgg taaaggcgac acaggaggag aaccgggagc tgaggagcag gtgtgaggag 2100
ctccacggga agaacctgga actggggaag atcatggaca ggttcgaaga ggttgtgtac 2160
caggccatgg aggaagttca gaagcagaag gaactttcca aagctgaaat ccagaaaagt 2220
ctaaaagaaa aagaccaact taccacagat ctgaactcca tggagaagtc cttctccgac 2280
ctcttcaagc gttttgagaa acagaaagag gtgatcgagg gctaccgcaa gaacgaagag 2340
tcagtgaaga agtgcggtga ggattacctg gcaaggatca cccaggaggg ccaggagtac 2400
caagccctga aggccacgc ggaggagaag ctgcagctgg caaacgagga gatcgccccag 2460
gtccggagca aggcccaggc ggaagcggtt gccctccagg ccagcctgag gaaggagcag 2520
atgcgcatcc agtcgctgga gaagacagtg gagcagaaga ctaaagagaa cgaggagctg 2580
accaggatct gcgacgacct catctccaag atggagaaga tctgacctcc acggagccgc 2640
tgtccccgcc cccctgctcc cgtctgtctg tcctgtctga ttctcttagg tgtcatgttc 2700
ttttttctgt cttgtcttca acttttttta aaactagatt gctttgaaaa catgactcaa 2760
taaaagtttc ctttcaattt aaaaaaaa
2788

```

&lt;210&gt; 455

&lt;211&gt; 720

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 455

```

Met Asp Asp Pro Lys Lys Glu Asp Ile Leu Leu Leu Ala Asp Glu Lys
1      5      10      15
Phe Asp Phe Asp Leu Ser Leu Ser Ser Ser Ala Asn Glu Asp Asp
20     25     30
Glu Val Phe Phe Gly Pro Phe Gly His Lys Glu Arg Cys Ile Ala Ala
35     40     45
Ser Leu Glu Leu Asn Asn Pro Val Pro Glu Gln Pro Pro Leu Pro Thr
50     55     60
Ser Glu Ser Pro Phe Ala Trp Ser Pro Leu Ala Gly Glu Lys Phe Val
65     70     75     80
Glu Val Tyr Lys Glu Ala His Leu Leu Ala Leu His Ile Glu Ser Ser
85     90     95
Ser Arg Asn Gln Ala Ala Gln Ala Ala Lys Pro Glu Asp Pro Arg Ser
100    105    110
Gln Gly Val Glu Arg Phe Ile Gln Glu Ser Lys Leu Lys Ile Asn Leu
115    120    125
Phe Glu Lys Glu Lys Glu Met Lys Lys Ser Pro Thr Ser Leu Lys Arg
130    135    140
Glu Thr Tyr Tyr Leu Ser Asp Ser Pro Leu Leu Gly Pro Pro Val Gly
145    150    155    160
Glu Pro Arg Leu Leu Ala Ser Ser Pro Ala Leu Pro Ser Ser Gly Ala
165    170    175
Gln Ala Arg Leu Thr Arg Ala Pro Gly Pro Pro His Ser Ala His Ala
180    185    190
Leu Pro Arg Glu Ser Cys Thr Ala His Ala Ala Ser Gln Ala Ala Thr
195    200    205
Gln Arg Lys Pro Gly Thr Lys Leu Leu Leu Pro Arg Ala Ala Ser Val
210    215    220
Arg Gly Arg Ser Ile Pro Gly Ala Ala Glu Lys Pro Lys Lys Glu Ile
225    230    235    240
Pro Ala Ser Pro Ser Arg Thr Lys Ile Pro Ala Glu Lys Glu Ser His
245    250    255
Arg Asp Val Leu Pro Asp Lys Pro Ala Pro Gly Ala Val Asn Val Pro
260    265    270
Ala Ala Gly Ser His Leu Gly Gln Gly Lys Arg Ala Ile Pro Val Pro
275    280    285
Asn Lys Leu Gly Leu Lys Lys Thr Leu Leu Lys Ala Pro Gly Ser Thr

```



290	295	300
Ser Asn Leu Ala Arg Lys Ser Ser Ser Gly Pro Val Trp Ser Gly Ala		
305	310	315
Ser Ser Ala Cys Thr Ser Pro Ala Val Gly Lys Ala Lys Ser Ser Glu		
	325	330
Phe Ala Ser Ile Pro Ala Asn Ser Ser Arg Pro Leu Ser Asn Ile Ser		
	340	345
Lys Ser Gly Arg Met Gly Pro Ala Met Leu Arg Pro Ala Leu Pro Ala		
	355	360
Gly Pro Val Gly Ala Ser Ser Trp Gln Ala Lys Arg Val Asp Val Ser		
	370	375
Glu Leu Ala Ala Glu Gln Leu Thr Ala Pro Pro Ser Ala Ser Pro Thr		
385	390	395
Gln Pro Gln Thr Pro Glu Gly Gly Gly Gln Trp Leu Asn Ser Ser Cys		
	405	410
Ala Trp Ser Glu Ser Ser Gln Leu Asn Lys Thr Arg Ser Ile Arg Arg		
	420	425
Arg Asp Ser Cys Leu Asn Ser Lys Thr Lys Val Met Pro Thr Pro Thr		
	435	440
Asn Gln Phe Lys Ile Pro Lys Phe Ser Ile Gly Asp Ser Pro Asp Ser		
	450	455
Ser Thr Pro Lys Leu Ser Arg Ala Gln Arg Pro Gln Ser Cys Thr Ser		
465	470	475
Val Gly Arg Val Thr Val His Ser Thr Pro Val Arg Arg Ser Ser Gly		
	485	490
Pro Ala Pro Gln Ser Leu Leu Ser Ala Arg Arg Val Ser Ala Leu Pro		
	500	505
Thr Pro Ala Ser Arg Arg Cys Ser Gly Leu Pro Pro Met Thr Pro Lys		
	515	520
Thr Met Pro Arg Ala Val Gly Ser Pro Leu Cys Val Pro Ala Arg Arg		
	530	535
Arg Ser Ser Glu Pro Arg Lys Asn Ser Ala Met Arg Thr Glu Pro Thr		
545	550	555
Arg Glu Ser Asn Arg Lys Thr Asp Ser Arg Leu Val Asp Val Ser Pro		
	565	570
Asp Arg Gly Ser Pro Pro Ser Arg Val Pro Gln Ala Leu Asn Phe Ser		
	580	585
Pro Glu Glu Ser Asp Ser Thr Phe Ser Lys Ser Thr Ala Thr Glu Val		
	595	600
Ala Arg Glu Glu Ala Lys Pro Gly Gly Asp Ala Ala Pro Ser Glu Ala		
	610	615
Leu Leu Val Asp Ile Lys Leu Glu Pro Leu Ala Val Thr Pro Asp Ala		
625	630	635
Ala Ser Gln Pro Leu Ile Asp Leu Pro Leu Ile Asp Phe Cys Asp Thr		
	645	650
Pro Glu Ala His Val Ala Val Gly Ser Glu Ser Arg Pro Leu Ile Asp		
	660	665
Leu Met Thr Asn Thr Pro Asp Met Asn Lys Asn Val Ala Lys Pro Ser		
	675	680
Pro Val Val Gly Gln Leu Ile Asp Leu Ser Ser Pro Leu Ile Gln Leu		
	690	695
Ser Pro Glu Ala Asp Lys Glu Asn Val Asp Ser Pro Leu Leu Lys Phe		
705	710	715
		720

<210> 456  
 <211> 2917  
 <212> DNA  
 <213> Homo sapiens

<400> 456

```

ggcacgaggg cggcggccgc gatgagcctt cagcctgccg ggcaggggac gtgaacatgg 60
atgaccctaa gaaggaagac attcttcttt tggccgatga aaaatttgac ttcgatcttt 120
cattgtcttc ttcgagtgc aatgaagatg atgaagtctt ctccggaccc tttggacata 180
aagaaagatg tattgctgcc agcttggaaat taaataatcc gggtcccgaa cagcctccgt 240
tgccacatc tgagagtccc ttgacctggc gccctctggc cggggagaag ttcgtggagg 300
tgtacaaaga agctcactta ctggctttac acattgagag cagcagccgg aaccaggcag 360
cccaagctgc caagcctgaa gaccctcgga gccaggcggt ggaaagattc atacaggagt 420
caaaattaaa aataaacctc tttgagaaag aaaaggaaat gaagaaaagc cccacgtctc 480
ttaaaggga gacatactac ctgtcagaca gcccttggct ggggccccct gtgggtgagc 540
ctcggctctt ggccctctcc cggccctgc ccagctctgg tgccagggc cgcctcacc 600
gggcgcggg gcctccgcac tctgtctatg ctttggccag ggaatcatgc actgctcatg 660
ctgcaagtca ggcagcgact cagaggaaag ccgggaccaa attgtgtctg cctcgagcgg 720
cctctgttag aggaagaagc atccctgggg ctgcggagaa gcccaagaaa gagattccag 780
ctagtccttc caggacaaaa atcccgctg agaaggaatc ccaccgggat gttctccctg 840
acaaacctgc ccgggtgtgt gtcaatgtgc cggccgcgg aagccacttg ggcagggca 900
agcgggcgat ccctgttcca aacaagttgg ggctgaagaa gaccctgtta aaagcaccgc 960
gctctaccag caatctcgca aggaagtcct cctcggggcc tgtttggagc ggggcatcca 1020
gtgctgacac atccccagca gtgggcaaa ctaaatacag tgaatttgca agtattcctg 1080
caaatagtct ccggcctctg tcaaacatca gacgtcagg cagaatggga cccgccatgc 1140
tgccggcagc tctgcctgca ggccctgtgg gggcactctc ctggcaggcc aagcgggtcg 1200
atgtttctga gctggcagcg gagcagctca cggcaccccc ctacagatcc cccaccaca 1260
cccagactcc ggaaggtggc ggccagtggc tgaactccag ttgcgcttgg tcagaatctt 1320
ctcaattgaa taagactaga agtatcagac ggcgagatcc ctgtctaaat tccaagaca 1380
aggttatgcc tactctaca aatcaattta aaattcctaa gttttctatt ggtgactccc 1440
cggacagctc aacaccaaa ctttcgcggg cagcgcggcc gcagtcgtgc acgtcagttg 1500
gcagggtcac tgtccacagc acccgggtta gacgtctatc tgggccagca ccacaaagcc 1560
tgctgagcgc acggcgtgtg tcagccttgc ccacaccgc cagccggcgc tgcctggcc 1620
ttccaccgat gacccccaaa acgatgcca gggccgtggg ctctccctg tgtgtgccag 1680
ctcggagacg ttctctgag ccccgcaaga actctgcaat gagaactgaa ccaacaaggg 1740
agagcaacag aaagacagat tccaggtctg tggatgtgtc ccctgacagg ggttctctc 1800
cttcccggtg gcctcaggca cttaactttt ctccagagga aagcgattct actttctcca 1860
aaagtactgc cacagaagta gctcgggagg aagccaagcc ggggtggagt gcagccccta 1920
gtgaggctct tctttagat atcaaaactg aaccactgc ggtcactcca gatgtgcaa 1980
gccagccctt cattgacctt cctctcatcg acttctgcga taccacagaa gcacacgtgg 2040
ctgtaggatc tgaagcagg cctctgatcg acctcatgac aaacactcca gacatgaata 2100
aaaatgtggc caaaccttca ccggtgggtg gacagctcat agacctgagc tccctctga 2160
tccagctgag cctgaggct gacaaggaga acgtggattc cccactctc aagttctaag 2220
ccgaacccaa tcttttgcc ttgaaagaaca gccctaaagt ggttttcaac cctcagaaac 2280
aagctttagg ctggtgcag tggcttacac ttgtaacctt agaacttggg aggtgaggt 2340
gggcggatta ctgagccca ggagttcggg accagcctgg gaaatatagt gaaactcctg 2400
tccctacaaa aagtacaaa attagccggg tgtggtagt catgcctgta gtcccagcta 2460
cttgggaggg tgaagtggga ggatggcctg agctcaagga gatgcaggct gcagtgagg 2520
gtgattgtgc cactgcactc cagcctgggc accaatgtga gaacctgtct tggaaaaaaa 2580
aaaaaaaaa aacatgtttt agtagaagtt ttatttgaaa aagaaaaata agcataaata 2640
tattcccagt gctggagagg gtgggctgag gactggggc cagcacggac caccacaggc 2700
ctctgcttcc cgcgccacc ctctcgtcg ccattctctg ggctggaatg tgaagcctca 2760
gtcactctaa atgaagaatt ttcttttgaa tgttttgtat gtaaaatagc aagtggctat 2820
ttttaaagtt aagtttgtat aaatagttag atattctaga tttacattaa attgtaaaat 2880
aaatggactt attgaagcat aaaaaaaaaa aaaaaaa 2917

```

<210> 457  
 <211> 389  
 <212> PRT  
 <213> Homo sapiens

<400> 457  
 Met Leu Ser Leu Arg Val Pro Leu Ala Pro Ile Thr Asp Pro Gln Gln  
 1 5 10 15  
 Leu Gln Leu Ser Pro Leu Lys Gly Leu Ser Leu Val Asp Lys Glu Asn  
 20 25 30  
 Thr Pro Pro Ala Leu Ser Gly Thr Arg Val Leu Ala Ser Lys Thr Ala

35	40	45
Arg Arg Ile Phe Gln Glu Pro Thr Glu Pro Lys Thr Lys Ala Ala Ala		
50	55	60
Pro Gly Val Glu Asp Glu Pro Leu Leu Arg Glu Asn Pro Arg Arg Phe		
65	70	75
Val Ile Phe Pro Ile Glu Tyr His Asp Ile Trp Gln Met Tyr Lys Lys		
85	90	95
Ala Glu Ala Ser Phe Trp Thr Ala Glu Val Asp Leu Ser Lys Asp		
100	105	110
Ile Gln His Trp Glu Ser Leu Lys Pro Glu Glu Arg Tyr Phe Ile Ser		
115	120	125
His Val Leu Ala Phe Phe Ala Ala Ser Asp Gly Ile Val Asn Glu Asn		
130	135	140
Leu Val Glu Arg Phe Ser Gln Glu Val Gln Ile Thr Glu Ala Arg Cys		
145	150	155
Phe Tyr Gly Phe Gln Ile Ala Met Glu Asn Ile His Ser Glu Met Tyr		
165	170	175
Ser Leu Leu Ile Asp Thr Tyr Ile Lys Asp Pro Lys Glu Arg Glu Phe		
180	185	190
Leu Phe Asn Ala Ile Glu Thr Met Pro Cys Val Lys Lys Lys Ala Asp		
195	200	205
Trp Ala Leu Arg Trp Ile Gly Asp Lys Glu Ala Thr Tyr Gly Glu Arg		
210	215	220
Val Val Ala Phe Ala Ala Val Glu Gly Ile Phe Phe Ser Gly Ser Phe		
225	230	235
Ala Ser Ile Phe Trp Leu Lys Lys Arg Gly Leu Met Pro Gly Leu Thr		
245	250	255
Phe Ser Asn Glu Leu Ile Ser Arg Asp Glu Gly Leu His Cys Asp Phe		
260	265	270
Ala Cys Leu Met Phe Lys His Leu Val His Lys Pro Ser Glu Glu Arg		
275	280	285
Val Arg Glu Ile Ile Ile Asn Ala Val Arg Ile Glu Gln Glu Phe Leu		
290	295	300
Thr Glu Ala Leu Pro Val Lys Leu Ile Gly Met Asn Cys Thr Leu Met		
305	310	315
Lys Gln Tyr Ile Glu Phe Val Ala Asp Arg Leu Met Leu Glu Leu Gly		
325	330	335
Phe Ser Lys Val Phe Arg Val Glu Asn Pro Phe Asp Phe Met Glu Asn		
340	345	350
Ile Ser Leu Glu Gly Lys Thr Asn Phe Phe Glu Lys Arg Val Gly Glu		
355	360	365
Tyr Gln Arg Met Gly Val Met Ser Ser Pro Thr Glu Asn Ser Phe Thr		
370	375	380
Leu Asp Ala Asp Phe		
385		

<210> 458  
 <211> 2500  
 <212> DNA  
 <213> Homo sapiens

<400> 458  
 cccaggcgca gccaatggga agggtcggag gcatggcaca gccaatggga agggccgggg 60  
 caccaaagcc aatgggaagg gccgggagcg cgcggcgcgg gagatttaaa ggctgctgga 120  
 gtgaggggtc gcccggtgcac cctgtcccag ccgtcctgtc ctggctgtc gctctgcttc 180  
 gctgcgcctc cactatgtct tccctccgtg tcccgtctgc gccatcacg gaccgcagc 240  
 agctgcagct ctgcgcgctg aaggggctca gcttggtcga caaggagaac acgccgccg 300  
 ccctgagcgg gaccgcgctc ctggccagca agaccgcgag gaggatcttc caggagccca 360  
 cggagccgaa aactaaagca gctgcccccg gcgtggagga tgagccgctg ctgagagaaa 420  
 acccccgcgg ctttgtcatc ttccccatcg agtaccatga tatctggcag atgtataaga 480

```

aggcagaggc ttccttttgg accgcccagg aggttgacct ctccaaggac attcagcact 540
gggaatccct gaaacccgag gagagatatt ttatatccca tgttctggct ttctttgcag 600
caagcgatgg catagtaaat gaaaacttgg tggagcgatt tagccaagaa gttcagatta 660
cagaagcccc ctgtttctat ggcttccaaa ttgccatgga aaacatacat tctgaaatgt 720
atagtcttct tattgacact gtcaagaaga aggagactg ggccttgccg tggattgggg 840
ccattgaaac gatgccttgt gaacgtgttg tagcctttgc tgcagtggaa ggcattttct 900
acaaagaggc tacctatggt atattctggc tcaagaaacg aggactgatg cctggcctca 960
tttccgggtc ttttgcgtcg agcagagatg agggtttaca ctgtgatttt gcttgccctga 1020
cattttctaa tgaacttatt aaaccatcgg aggagagagt aagagaaata attatcaatg 1080
tgttcaaaca cctgggtacac ttccctcactg aggccttgcc tgtgaagctc attgggatga 1140
ctgttcggat agaacaggag tacattgagt ttgtggcaga cagacttatg ctggaactgg 1200
attgcactct aatgaagcaa gttagagaacc catttgactt tatggagaat atttcaactg 1260
gttttagcaa ggttttcaga gtagagaacc catttgactt tatggagaat atttcaactg 1260
aaggaaagac taacttcttt gagaagagag taggcgagta tcagaggatg ggagtgatgt 1320
caagtccaac agagaattct tttaccttgg atgctgactt ctaaataaac tgaagatgtg 1380
cccttacttg gctgattttt tttttccatc tcataagaaa aatcagctga agtggtacca 1440
actagccaca ccattgaattg tccgtaattg tcattaacag catctttaaa actgtgtagc 1500
tacctcacia ccagtcctgt ctgttttatag tgctggtagt atcacctttt gccagaaggc 1560
ctggctggct gtgacttacc atagcagtga caatggcagt ctggccttta aagtggaggg 1620
tgacccttta gtgagcttag cacagcggga ttaaaccagc ctttaaccag cacagccagt 1680
taaaagatgc agcctcactg cttcaacgca gattttaatg tttactttaa tataaacctg 1740
gcactttaca aacaaataaa cattgttttg tactcacggc ggcgataata gcttgattta 1800
tttggtttct acaccaaata cattctcctg accactaatg ggagccaatt cacaattcac 1860
taagtgacta aagtaagtta aacttgtgta gactaagcat gtaattttta agttttattt 1920
taatgaatta aaatatgtgt taaccaactt taaagtcagt cctgtgtata cctagatatt 1980
agtcagttgg tgccagatag aagacaggtt gtgtttttat cctgtggcct gtgtagtgtc 2040
ctgggattct ctgccccctc tgagtagagt gttgtgggat aaaggaatct ctcagggcaa 2100
ggagcttctt aagttaaatc actagaaatt taggggtgat ctgggccttc atatgtgtga 2160
gaagccggtt cattttattt ctcactgtat tttcctcaac gtctggttga tgagaaaaaa 2220
ttcttgaaga gttttcatat gtgggagcta aggtagtatt gtaaaatttc aagtcacctc 2280
taaacaaaat gatccaccta agatcttgcc cctgttaagt ggtgaaatca actagaggtg 2340
gttcctacaa gttgttcatt ctagttttgt ttgggtgaag taggttgtgt gagttaattc 2400
atttatattt actatgtctg ttaaatacaga aattttttat tatctatgtt cttctagatt 2460
ttacctgtag ttcataaaaa aaaaaaaaaa aaaaaaaaaa 2500

```

<210> 459  
 <211> 890  
 <212> PRT  
 <213> Homo sapiens

```

<400> 459
Met Ser Gln Gly Ile Leu Ser Pro Pro Ala Gly Leu Leu Ser Asp Asp
1      5      10
Asp Val Val Val Ser Pro Met Phe Glu Ser Thr Ala Ala Asp Leu Gly
20     25     30
Ser Val Val Arg Lys Asn Leu Leu Ser Asp Cys Ser Val Val Ser Thr
35     40     45
Ser Leu Glu Asp Lys Gln Gln Val Pro Ser Glu Asp Ser Met Glu Lys
50     55     60
Val Lys Val Tyr Leu Arg Val Arg Pro Leu Leu Pro Ser Glu Leu Glu
65     70     75     80
Arg Gln Glu Asp Gln Gly Cys Val Arg Ile Glu Asn Val Glu Thr Leu
85     90     95
Val Leu Gln Ala Pro Lys Asp Ser Phe Ala Leu Lys Ser Asn Glu Arg
100    105    110
Gly Ile Gly Gln Ala Thr His Arg Phe Thr Phe Ser Gln Ile Phe Gly
115    120    125
Pro Glu Val Gly Gln Ala Ser Phe Phe Asn Leu Thr Val Lys Glu Met
130    135    140
Val Lys Asp Val Leu Lys Gly Gln Asn Trp Leu Ile Tyr Thr Tyr Gly
145    150    155    160

```

Val	Thr	Asn	Ser	Gly	Lys	Thr	His	Thr	Ile	Gln	Gly	Thr	Ile	Lys	Asp
				165					170					175	
Gly	Gly	Ile	Leu	Pro	Arg	Ser	Leu	Ala	Leu	Ile	Phe	Asn	Ser	Leu	Gln
			180					185					190		
Gly	Gln	Leu	His	Pro	Thr	Pro	Asp	Leu	Lys	Pro	Leu	Leu	Ser	Asn	Glu
		195					200					205			
Val	Ile	Trp	Leu	Asp	Ser	Lys	Gln	Ile	Arg	Gln	Glu	Glu	Met	Lys	Lys
	210					215					220				
Leu	Ser	Leu	Leu	Asn	Gly	Gly	Leu	Gln	Glu	Glu	Glu	Leu	Ser	Thr	Ser
225					230					235					240
Leu	Lys	Arg	Ser	Val	Tyr	Ile	Glu	Ser	Arg	Ile	Gly	Thr	Ser	Thr	Ser
				245					250						255
Phe	Asp	Ser	Gly	Ile	Ala	Gly	Leu	Ser	Ser	Ile	Ser	Gln	Cys	Thr	Ser
			260					265					270		
Ser	Ser	Gln	Leu	Asp	Glu	Thr	Ser	His	Arg	Trp	Ala	Gln	Pro	Asp	Thr
		275						280				285			
Ala	Pro	Leu	Pro	Val	Pro	Ala	Asn	Ile	Arg	Phe	Ser	Ile	Trp	Ile	Ser
	290					295					300				
Phe	Phe	Glu	Ile	Tyr	Asn	Glu	Leu	Leu	Tyr	Asp	Leu	Leu	Glu	Pro	Pro
305					310					315					320
Ser	Gln	Gln	Arg	Lys	Arg	Gln	Thr	Leu	Arg	Leu	Cys	Glu	Asp	Gln	Asn
				325					330						335
Gly	Asn	Pro	Tyr	Val	Lys	Asp	Leu	Asn	Trp	Ile	His	Val	Gln	Asp	Ala
			340					345					350		
Glu	Glu	Ala	Trp	Lys	Leu	Leu	Lys	Val	Gly	Arg	Lys	Asn	Gln	Ser	Phe
		355					360					365			
Ala	Ser	Thr	His	Leu	Asn	Gln	Asn	Ser	Ser	Arg	Ser	His	Ser	Ile	Phe
	370					375					380				
Ser	Ile	Arg	Ile	Leu	His	Leu	Gln	Gly	Glu	Gly	Asp	Ile	Val	Pro	Lys
385					390					395					400
Ile	Ser	Glu	Leu	Ser	Leu	Cys	Asp	Leu	Ala	Gly	Ser	Glu	Arg	Cys	Lys
				405					410					415	
Asp	Gln	Lys	Ser	Gly	Glu	Arg	Leu	Lys	Glu	Ala	Gly	Asn	Ile	Asn	Thr
			420					425					430		
Ser	Leu	His	Thr	Leu	Gly	Arg	Cys	Ile	Ala	Ala	Leu	Arg	Gln	Asn	Gln
		435					440					445			
Gln	Asn	Arg	Ser	Lys	Gln	Asn	Leu	Val	Pro	Phe	Arg	Asp	Ser	Lys	Leu
	450					455					460				
Thr	Arg	Val	Phe	Gln	Gly	Phe	Phe	Thr	Gly	Arg	Gly	Arg	Ser	Cys	Met
465					470					475					480
Ile	Val	Asn	Val	Asn	Pro	Cys	Ala	Ser	Thr	Tyr	Asp	Glu	Thr	Leu	His
				485					490					495	
Val	Ala	Lys	Phe	Ser	Ala	Ile	Ala	Ser	Gln	Leu	Val	His	Ala	Pro	Pro
			500					505					510		
Met	Gln	Leu	Gly	Phe	Pro	Ser	Leu	His	Ser	Phe	Ile	Lys	Glu	His	Ser
		515					520					525			
Leu	Gln	Val	Ser	Pro	Ser	Leu	Glu	Lys	Gly	Ala	Lys	Ala	Asp	Thr	Gly
		530				535					540				
Leu	Asp	Asp	Asp	Ile	Glu	Asn	Glu	Ala	Asp	Ile	Ser	Met	Tyr	Gly	Lys
545					550					555					560
Glu	Glu	Leu	Leu	Gln	Val	Val	Glu	Ala	Met	Lys	Thr	Leu	Leu	Leu	Lys
				565					570					575	
Glu	Arg	Gln	Glu	Lys	Leu	Gln	Leu	Glu	Met	His	Leu	Arg	Asp	Glu	Ile
			580					585					590		
Cys	Asn	Glu	Met	Val	Glu	Gln	Met	Gln	Gln	Arg	Glu	Gln	Trp	Cys	Ser
		595					600					605			
Glu	His	Leu	Asp	Thr	Gln	Lys	Glu	Leu	Leu	Glu	Glu	Met	Tyr	Glu	Glu
	610					615					620				
Lys	Leu	Asn	Ile	Leu	Lys	Glu	Ser	Leu	Thr	Ser	Phe	Tyr	Gln	Glu	Glu
625					630					635					640
Ile	Gln	Glu	Arg	Asp	Glu	Lys	Ile	Glu	Glu	Leu	Glu	Ala	Leu	Leu	Gln

BNSDOCID: <WO 2004055050A2 | >

```

acctctctac acaccctggg ccgctgtatt gctgcccttc gtcaaaacca gcagaaccgg 1380
tcaaagcaga acctgggttcc cttccgtgac agcaagttga ctcgagtgtt ccaaggtttc 1440
ttcacaggcc gaggccgttc ctgcatgatt gtcaatgtga atccctgtgc atctacctat 1500
gatgaaactc ttcattgtggc caagttctca gccattgcta gccagcttgt gcatgcccc 1560
cctatgcaac tgggattccc atccctgcac tcgttcacatc aggaacatag tcttcaggta 1620
tccccagct tagagaaagg ggctaaggca gacacaggcc ttgatgatga tattgaaaat 1680
gaagctgaca tctccatgta tggcaaagag gagctcctac aagttgtgga agccatgaag 1740
acactgcttt tgaaggaacg accagaaaag ctacagctgg agatgcatct ccgagatgaa 1800
atttgcaatg agatggtaga acagatgcaa cagcgggaac agtggtgcag tgaacatttg 1860
gacacccaaa aggaactatt ggaggaaatg tatgaagaaa aactaaatat cctcaaggag 1920
tactgacaa gtttttacc aagaagagatt caggagcggg atgaaaagat tgaagagcta 1980
gaagctctct tgcaggaagc cagacaacag tcagtggccc atcagcaatc agggctctgaa 2040
ttggccctac ggcggtcaca aaggttggca gcttctgcct ccaccagca gcttcaggag 2100
gttaaagcta aattacagca gtgcaaagca gagctaaact ctaccactga agagttgcat 2160
aagtatcaga aaatgttaga accaccacc cttcaccat tgatgtggac 2220
aagaagttag aagagggcca gaagaatata aggtgttgc ggacagagct tcagaactt 2280
ggtgagctc tccaatcagc agagagagct tgttgccaca gcaactggggc aggaaaactt 2340
cgtcaagcct tgaccacttg tgatgacatc ttaatcaaac aggaccagac tctggctgaa 2400
ctgcagaaca acatgggtgct agtgaaactg gaccttcgga agaaggcagc atgtattgct 2460
gagcagtatc atactgtgtt gaaactccaa ggccagggtt ctgccccaaa ggccttgggt 2520
accaaccagg aaaactcagca accaaaccaa caaccaccag ggaagaaacc attccttcga 2580
aatttacttc cccgaacacc aacctgccaa agctcaacag actgcagccc ttatgcccgg 2640
atcctacgct cagggcggtc cctttactc aaatctgggc cttttggcaa aaagtactaa 2700
ggctgtgggg aaagagaaga gcagtcattg ccctgaggtg ggctcagctac tctcctgaag 2760
aaataggtct cttttatgct ttaccatata tcaggaatta tatccaggat gcaatactca 2820
gacactagct tttttctcac ttttgtatta taaccaccta tgtaatctca tgttgttgtt 2880
tttttttatt tacttatatg atttctatgc acacaaaaac agttatatta aagatattat 2940
tggtcacatt ttttattgaa aaaaaaaaaa aa 2972

```

<210> 461  
 <211> 719  
 <212> PRT  
 <213> Homo sapiens

<400> 461

Met	Ala	Leu	Lys	Asp	Tyr	Ala	Leu	Glu	Lys	Glu	Lys	Val	Lys	Lys	Phe
1			5					10						15	
Leu	Gln	Glu	Phe	Tyr	Gln	Asp	Asp	Glu	Leu	Gly	Lys	Lys	Gln	Phe	Lys
		20						25					30		
Tyr	Gly	Asn	Gln	Leu	Val	Arg	Leu	Ala	His	Arg	Glu	Gln	Val	Ala	Leu
		35					40					45			
Tyr	Val	Asp	Leu	Asp	Asp	Val	Ala	Glu	Asp	Asp	Pro	Glu	Leu	Val	Asp
	50					55					60				
Ser	Ile	Cys	Glu	Asn	Ala	Arg	Arg	Tyr	Ala	Lys	Leu	Phe	Ala	Asp	Ala
	65				70				75					80	
Val	Gln	Glu	Leu	Leu	Pro	Gln	Tyr	Lys	Glu	Arg	Glu	Val	Val	Asn	Lys
			85					90						95	
Asp	Val	Leu	Asp	Val	Tyr	Ile	Glu	His	Arg	Leu	Met	Met	Glu	Gln	Arg
		100					105						110		
Ser	Arg	Asp	Pro	Gly	Met	Val	Arg	Ser	Pro	Gln	Asn	Gln	Tyr	Pro	Ala
		115					120					125			
Glu	Leu	Met	Arg	Arg	Phe	Glu	Leu	Tyr	Phe	Gln	Gly	Pro	Ser	Ser	Asn
	130					135					140				
Lys	Pro	Arg	Val	Ile	Arg	Glu	Val	Arg	Ala	Asp	Ser	Val	Gly	Lys	Leu
	145				150				155					160	
Val	Thr	Val	Arg	Gly	Ile	Val	Thr	Arg	Val	Ser	Glu	Val	Lys	Pro	Lys
			165				170							175	
Met	Val	Val	Ala	Thr	Tyr	Thr	Cys	Asp	Gln	Cys	Gly	Ala	Glu	Thr	Tyr
		180					185						190		
Gln	Pro	Ile	Gln	Ser	Pro	Thr	Phe	Met	Pro	Leu	Ile	Met	Cys	Pro	Ser
	195						200					205			

Gln Glu Cys Gln Thr Asn Arg Ser Gly Gly Arg Leu Tyr Leu Gln Thr  
 210 215 220  
 Arg Gly Ser Arg Phe Ile Lys Phe Gln Glu Met Lys Met Gln Glu His  
 225 230 235 240  
 Ser Asp Gln Val Pro Val Gly Asn Ile Pro Arg Ser Ile Thr Val Leu  
 245 250 255  
 Val Glu Gly Glu Asn Thr Arg Ile Ala Gln Pro Gly Asp His Val Ser  
 260 265 270  
 Val Thr Gly Ile Phe Leu Pro Ile Leu Arg Thr Gly Phe Arg Gln Val  
 275 280 285  
 Val Gln Gly Leu Leu Ser Glu Thr Tyr Leu Glu Ala His Arg Ile Val  
 290 295 300  
 Lys Met Asn Lys Ser Glu Asp Asp Glu Ser Gly Ala Gly Glu Leu Thr  
 305 310 315 320  
 Arg Glu Glu Leu Arg Gln Ile Ala Glu Glu Asp Phe Tyr Glu Lys Leu  
 325 330 335  
 Ala Ala Ser Ile Ala Pro Glu Ile Tyr Gly His Glu Asp Val Lys Lys  
 340 345 350  
 Ala Leu Leu Leu Leu Val Gly Gly Val Asp Gln Ser Pro Arg Gly  
 355 360 365  
 Met Lys Ile Arg Gly Asn Ile Asn Ile Cys Leu Met Gly Asp Pro Gly  
 370 375 380  
 Val Ala Lys Ser Gln Leu Ser Tyr Ile Asp Arg Leu Ala Pro Arg  
 385 390 395 400  
 Ser Gln Tyr Thr Thr Gly Arg Gly Ser Ser Gly Val Gly Leu Thr Ala  
 405 410 415  
 Ala Val Leu Arg Asp Ser Val Ser Gly Glu Leu Thr Leu Glu Gly Gly  
 420 425 430  
 Ala Leu Val Leu Ala Asp Gln Gly Val Cys Cys Ile Asp Glu Phe Asp  
 435 440 445  
 Lys Met Ala Glu Ala Asp Arg Thr Ala Ile His Glu Val Met Glu Gln  
 450 455 460  
 Gln Thr Ile Ser Ile Ala Lys Ala Gly Ile Leu Thr Thr Leu Asn Ala  
 465 470 475 480  
 Arg Cys Ser Ile Leu Ala Ala Ala Asn Pro Ala Tyr Gly Arg Tyr Asn  
 485 490 495  
 Pro Arg Arg Ser Leu Glu Gln Asn Ile Gln Leu Pro Ala Ala Leu Leu  
 500 505 510  
 Ser Arg Phe Asp Leu Leu Trp Leu Ile Gln Asp Arg Pro Asp Arg Asp  
 515 520 525  
 Asn Asp Leu Arg Leu Ala Gln His Ile Thr Tyr Val His Gln His Ser  
 530 535 540  
 Arg Gln Pro Pro Ser Gln Phe Glu Pro Leu Asp Met Lys Leu Met Arg  
 545 550 555 560  
 Arg Tyr Ile Ala Met Cys Arg Glu Lys Gln Pro Met Val Pro Glu Ser  
 565 570 575  
 Leu Ala Asp Tyr Ile Thr Ala Ala Tyr Val Glu Met Arg Arg Glu Ala  
 580 585 590  
 Trp Ala Ser Lys Asp Ala Thr Tyr Thr Ser Ala Arg Thr Leu Leu Ala  
 595 600 605  
 Ile Leu Arg Leu Ser Thr Ala Leu Ala Arg Leu Arg Met Val Asp Val  
 610 615 620  
 Val Glu Lys Glu Asp Val Asn Glu Ala Ile Arg Leu Met Glu Met Ser  
 625 630 635 640  
 Lys Asp Ser Leu Leu Gly Asp Lys Gly Gln Thr Ala Arg Thr Gln Arg  
 645 650 655  
 Pro Ala Asp Val Ile Phe Ala Thr Val Arg Glu Leu Val Ser Gly Gly  
 660 665 670  
 Arg Ser Val Arg Phe Ser Glu Ala Glu Gln Arg Cys Val Ser Arg Gly  
 675 680 685  
 Phe Thr Pro Ala Gln Phe Gln Ala Ala Leu Asp Glu Tyr Glu Glu Leu



690 695 700  
 Asn Val Trp Gln Val Asn Ala Ser Arg Thr Arg Ile Thr Phe Val  
 705 710 715

<210> 462  
 <211> 2821  
 <212> DNA  
 <213> Homo sapiens

<400> 462  
 cgcccccttc cagccccaag ggtctaggat acagtctttg tagatgagcg ggtccccctt 60  
 ggaggacaga atgaagaatt gggaaatcat ggccgttctg gagagtagac aagaagacgg 120  
 cgaaagtcgg gcctgccccg ccctgcgggc ccggaacaaa agaacgcgtg tgcgctggcc 180  
 cttaagagc gattctcttc cgcccgcgcc agctcgacc gcgggaaacc cggcgccctgc 240  
 actaccccg cggagattc ccttcgcacg ccgcacccg ctccccgtca ctcatcttag 300  
 gcccgcacgg tgattggctt gcggtagcgg ggaggtgaag aaggccgcct tgtccgattg 360  
 gcccgcacgc agtggcgccg gtcacgtggg gggcgacgtt tcgcgccaat ttcggttggc 420  
 cggccacagt ccaccgcgcg gagattctca gcttccccag gagcaagacc tctgagcccg 480  
 ccaagcgcg ccgcacggcc ctccggcagc atggcactga aggactacgc gctagagaag 540  
 gaaaagggtta agaagttctt acaagagttc taccaggatg atgaactcgg gaagaagcag 600  
 ttcaagtatg ggaaccagtt ggttcggctg gctcatcggg aacaggtggc tctgtatgtg 660  
 gacctggacg acgtagccga ggtgacccc agcttgggtg actcaatttg tgagaatgcc 720  
 aggcgtacg cgaagctctt tgctgatgcc gtacaagagc tgctgcctca gtacaaggag 780  
 agggagtg taaataaaga tgtcctggac gtttacattg agcatcggct aatgatggag 840  
 cagcggagtc gggaccctgg gatggctcga agccccaga accagtacc tgtgaactc 900  
 atgctcagat ttgagctgta ttttcaaggc cctagcagca acaagcctcg tgtgatccgg 960  
 gaagtgcggg ctgactctgt ggggaagttg gtaactgtgc gtggaatcgt cactcgtgtc 1020  
 tctgaagtca aacccaagat ggtggtggcc acttacactt gtgaccagtg tggggcagag 1080  
 acctaccagc cgatccagtc tcccactttc atgcctctga tcatgtgcc aagccaggag 1140  
 tgccaaacca accgctcagg agggcggtg tatctgcaga cacggggctc cagattcattc 1200  
 aaattccagg agatgaagat gcaagaacat agtgatcagg tgccctgtgg aaatatccct 1260  
 cgtagtatca cggctgctgtt agaaggagag aacacaagga ttgccagcc tggagaccac 1320  
 gtacgcgtca ctggtatttt cttgccaatc ctgcgcactg ggttccgaca ggtggtacag 1380  
 ggtttactct cagaaaccta cctggaagcc catcggattg tgaagatgaa caagagttag 1440  
 gatgatgagt ctggggctgg agagctcacc agggaggagc tgaggcaaat tgcagaggag 1500  
 gattttctacg aaaagctggc agcttcaatc gccccagaaa tatacgggca tgaagatgtg 1560  
 aagaaggcac tgcgtctcct gctagtcggg ggtgtggacc agtctcctcg aggcattgaaa 1620  
 atccggggca acatcaacat ctgtctgatg ggggatcctg gtgtggccaa gtctcagctc 1680  
 ctgtcataca ttgatcgact ggcgcctcgc agccagtaca caacaggccg gggctcctca 1740  
 ggagtggggc ttacggcagc tgtgctgaga gactccgtga gtggagaact gaccttagag 1800  
 ggtggggccc tgggtgctggc tgaccagggt gtgtgctgca ttgatgagtt cgacaagatg 1860  
 gctgaggccg accgcacagc catccacgag gtcattggag agcagaccat ctccattgcc 1920  
 aaggccggca ttctcaccac actcaatgcc cgctgctcca tcctggctgc cgccaacct 1980  
 gcctacgggc gctacaaccc tcgccgcagc ctggagcaga acatacagct acctgctgca 2040  
 ctgctctccc ggtttgacct cctctggctg attcaggacc ggcccagccg agacaatgac 2100  
 ctacggttg cccagcacat cacctatgtg caccagcaca gccggcagcc cccctcccag 2160  
 tttgaacctc tggacatgaa gctcatgagg cgttacatag ccatgtgccg cgagaagcag 2220  
 cccatggtgc cagagtctct ggtgactac atcacagcag catacgtgga gatgaggcga 2280  
 gaggtttggg ctagtaagga tgccacctat acttctgccc ggacctgct ggctatcctg 2340  
 cgcttttcca ctgctctggc acgtctgaga atggtggatg tgggtggaga agaagatgtg 2400  
 aatgaagcca tcaggcta at ggagatgtca aaggactctc ttctaggaga caaggggcag 2460  
 acagctagga ctacagagacc agcagatgtg atatttgcca ccgtccgtga actggtctca 2520  
 gggggccgaa gtgtccggtt ctctgaggca gagcagcgt gtgtatctcg tggcttcaca 2580  
 cccgccaggt tccaggcggc tctggatgaa tatgaggagc tcaatgtctg gcaggtcaat 2640  
 gcttcccga caggatcac tttgtctga ttccagcctg cttgcaaccc tggggtcctc 2700  
 ttgttccctg cctggcctgc ccttgggaag ggcagtgat gcctttgagg ggaaggagga 2760  
 gccccctttt ctcccatgct gcacttactc cttttgctaa taaaagtgtt tgtagattgt 2820  
 c 2821

<210> 463

<211> 1124  
 <212> PRT  
 <213> Homo sapiens

<400> 463

```

Met Asp Pro Phe Thr Glu Lys Leu Leu Glu Arg Thr Arg Ala Arg Arg
 1          5          10          15
Glu Asn Leu Gln Arg Lys Met Ala Glu Arg Pro Thr Ala Ala Pro Arg
 20          25          30
Ser Met Thr His Ala Lys Arg Ala Arg Gln Pro Leu Ser Glu Ala Ser
 35          40          45
Asn Gln Gln Pro Leu Ser Gly Gly Glu Glu Lys Ser Cys Thr Lys Pro
 50          55          60
Ser Pro Ser Lys Lys Arg Cys Ser Asp Asn Thr Glu Val Glu Val Ser
 65          70          75          80
Asn Leu Glu Asn Lys Gln Pro Val Glu Ser Thr Ser Ala Lys Ser Cys
 85          90          95
Ser Pro Ser Pro Val Ser Pro Gln Val Gln Pro Gln Ala Ala Asp Thr
100          105          110
Ile Ser Asp Ser Val Ala Val Pro Ala Ser Leu Leu Gly Met Arg Arg
115          120          125
Gly Leu Asn Ser Arg Leu Glu Ala Thr Ala Ala Ser Ser Val Lys Thr
130          135          140
Arg Met Gln Lys Leu Ala Glu Gln Arg Arg Arg Trp Asp Asn Asp Asp
145          150          155          160
Met Thr Asp Asp Ile Pro Glu Ser Ser Leu Phe Ser Pro Met Pro Ser
165          170          175
Glu Glu Lys Ala Ala Ser Pro Pro Arg Pro Leu Leu Ser Asn Ala Ser
180          185          190
Ala Thr Pro Val Gly Arg Arg Gly Arg Leu Ala Asn Leu Ala Ala Thr
195          200          205
Ile Cys Ser Trp Glu Asp Asp Val Asn His Ser Phe Ala Lys Gln Asn
210          215          220
Ser Val Gln Glu Gln Pro Gly Thr Ala Cys Leu Ser Lys Phe Ser Ser
225          230          235          240
Ala Ser Gly Ala Ser Ala Arg Ile Asn Ser Ser Ser Val Lys Gln Glu
245          250          255
Ala Thr Phe Cys Ser Gln Arg Asp Gly Asp Ala Ser Leu Asn Lys Ala
260          265          270
Leu Ser Ser Ser Ala Asp Asp Ala Ser Leu Val Asn Ala Ser Ile Ser
275          280          285
Ser Ser Val Lys Ala Thr Ser Pro Val Lys Ser Thr Thr Ser Ile Thr
290          295          300
Asp Ala Lys Ser Cys Glu Gly Gln Asn Pro Glu Leu Leu Pro Lys Thr
305          310          315          320
Pro Ile Ser Pro Leu Lys Thr Gly Val Ser Lys Pro Ile Val Lys Ser
325          330          335
Thr Leu Ser Gln Thr Val Pro Ser Lys Gly Glu Leu Ser Arg Glu Ile
340          345          350
Cys Leu Gln Ser Gln Ser Lys Asp Lys Ser Thr Thr Pro Gly Gly Thr
355          360          365
Gly Ile Lys Pro Phe Leu Glu Arg Phe Gly Glu Arg Cys Gln Glu His
370          375          380
Ser Lys Glu Ser Pro Ala Arg Ser Thr Pro His Arg Thr Pro Ile Ile
385          390          395          400
Thr Pro Asn Thr Lys Ala Ile Gln Glu Arg Leu Phe Lys Gln Asp Thr
405          410          415
Ser Ser Ser Thr Thr His Leu Ala Gln Gln Leu Lys Gln Glu Arg Gln
420          425          430
Lys Glu Leu Ala Cys Leu Arg Gly Arg Phe Asp Lys Gly Asn Ile Trp
435          440          445

```

Ser Ala Glu Lys Gly Gly Asn Ser Lys Ser Lys Gln Leu Glu Thr Lys  
 450 455 460  
 Gln Glu Thr His Cys Gln Ser Thr Pro Leu Lys Lys His Gln Gly Val  
 465 470 475 480  
 Ser Lys Thr Gln Ser Leu Pro Val Thr Glu Lys Val Thr Glu Asn Gln  
 485 490 495  
 Ile Pro Ala Lys Asn Ser Ser Thr Glu Pro Lys Gly Phe Thr Glu Cys  
 500 505 510  
 Glu Met Thr Lys Ser Ser Pro Leu Lys Ile Thr Leu Phe Leu Glu Glu  
 515 520 525  
 Asp Lys Ser Leu Lys Val Thr Ser Asp Pro Lys Val Glu Gln Lys Ile  
 530 535 540  
 Glu Val Ile Arg Glu Ile Glu Met Ser Val Asp Asp Asp Asp Ile Asn  
 545 550 555 560  
 Ser Ser Lys Val Ile Asn Asp Leu Phe Ser Asp Val Leu Glu Glu Gly  
 565 570 575  
 Glu Leu Asp Met Glu Lys Ser Gln Glu Glu Met Asp Gln Ala Leu Ala  
 580 585 590  
 Glu Ser Ser Glu Glu Gln Glu Asp Ala Leu Asn Ile Ser Ser Met Ser  
 595 600 605  
 Leu Leu Ala Pro Leu Ala Gln Thr Val Gly Val Val Ser Pro Glu Ser  
 610 615 620  
 Leu Val Ser Thr Pro Arg Leu Glu Leu Lys Asp Thr Ser Arg Ser Asp  
 625 630 635 640  
 Glu Ser Pro Lys Pro Gly Lys Phe Gln Arg Thr Arg Val Pro Arg Ala  
 645 650 655  
 Glu Ser Gly Asp Ser Leu Gly Ser Glu Asp Arg Asp Leu Leu Tyr Ser  
 660 665 670  
 Ile Asp Ala Tyr Arg Ser Gln Arg Phe Lys Glu Thr Glu Arg Pro Ser  
 675 680 685  
 Ile Lys Gln Val Ile Val Arg Lys Glu Asp Val Thr Ser Lys Leu Asp  
 690 695 700  
 Glu Lys Asn Asn Ala Phe Pro Cys Gln Val Asn Ile Lys Gln Lys Met  
 705 710 715 720  
 Gln Glu Leu Asn Asn Glu Ile Asn Met Gln Gln Thr Val Ile Tyr Gln  
 725 730 735  
 Ala Ser Gln Ala Leu Asn Cys Cys Val Asp Glu Glu His Gly Lys Gly  
 740 745 750  
 Ser Leu Glu Glu Ala Glu Ala Glu Arg Leu Leu Leu Ile Ala Thr Gly  
 755 760 765  
 Lys Arg Thr Leu Leu Ile Asp Glu Leu Asn Lys Leu Lys Asn Glu Gly  
 770 775 780  
 Pro Gln Arg Lys Asn Lys Ala Ser Pro Gln Ser Glu Phe Met Pro Ser  
 785 790 795 800  
 Lys Gly Ser Val Thr Leu Ser Glu Ile Arg Leu Pro Leu Lys Ala Asp  
 805 810 815  
 Phe Val Cys Ser Thr Val Gln Lys Pro Asp Ala Ala Asn Tyr Tyr Tyr  
 820 825 830  
 Leu Ile Ile Leu Lys Ala Gly Ala Glu Asn Met Val Ala Thr Pro Leu  
 835 840 845  
 Ala Ser Thr Ser Asn Ser Leu Asn Gly Asp Ala Leu Thr Phe Thr Thr  
 850 855 860  
 Thr Phe Thr Leu Gln Asp Val Ser Asn Asp Phe Glu Ile Asn Ile Glu  
 865 870 875 880  
 Val Tyr Ser Leu Val Gln Lys Lys Asp Pro Ser Gly Leu Asp Lys Lys  
 885 890 895  
 Lys Lys Thr Ser Lys Ser Lys Ala Ile Thr Pro Lys Arg Leu Leu Thr  
 900 905 910  
 Ser Ile Thr Thr Lys Ser Asn Ile His Ser Ser Val Met Ala Ser Pro  
 915 920 925  
 Gly Gly Leu Ser Ala Val Arg Thr Ser Asn Phe Ala Leu Val Gly Ser

930	935	940
Tyr Thr Leu Ser Leu Ser Ser Val Gly Asn Thr Lys Phe Val Leu Asp		
945	950	955
Lys Val Pro Phe Leu Ser Ser Leu Glu Gly His Ile Tyr Leu Lys Ile		960
	965	970
Lys Cys Gln Val Asn Ser Ser Val Glu Glu Arg Gly Phe Leu Thr Ile		975
	980	985
Phe Glu Asp Val Ser Gly Phe Gly Ala Trp His Arg Arg Trp Cys Val		990
	995	1000
Leu Ser Gly Asn Cys Ile Ser Tyr Trp Thr Tyr Pro Asp Asp Glu Lys		1005
	1010	1015
Arg Lys Asn Pro Ile Gly Arg Ile Asn Leu Ala Asn Cys Thr Ser Arg		1020
1025	1030	1035
Gln Ile Glu Pro Ala Asn Arg Glu Phe Cys Ala Arg Arg Asn Thr Phe		1040
	1045	1050
Glu Leu Ile Thr Val Arg Pro Gln Arg Glu Asp Asp Arg Glu Thr Leu		1055
	1060	1065
Val Ser Gln Cys Arg Asp Thr Leu Cys Val Thr Lys Asn Trp Leu Ser		1070
	1075	1080
Ala Asp Thr Lys Glu Glu Arg Asp Leu Trp Met Gln Lys Leu Asn Gln		1085
	1090	1095
Val Leu Val Asp Ile Arg Leu Trp Gln Pro Asp Ala Cys Tyr Lys Pro		1100
1105	1110	1115
Ile Gly Lys Pro		1120

<210> 464  
 <211> 4786  
 <212> DNA  
 <213> Homo sapiens

<400> 464

ctcggcgctg	aaattcaaat	ttgaacggct	gcagaggccg	agtcgcgtcac	tggaagccga	60
gaggagagga	cagctgggtg	tgggagagtt	ccccgcctc	agactcctgg	ttttttccag	120
gagacacact	gagctgagac	tcacttttct	cttcctgaat	ttgaaccacc	gtttccatcg	180
tctcgtagtc	cgacgcctgg	ggcgatggat	ccgtttacgg	agaaactgct	ggagcgaacc	240
cgtgccaggg	gagagaatct	tcagagaaaa	atggctgaga	ggcccacagc	agctccaagg	300
tctatgactc	atgctaagcg	agctagacag	ccactttcag	aagcaagtaa	ccagcagccc	360
ctctctgggtg	gtgaagagaa	atcttgtaca	aaaccatcgc	catcaaaaaa	acgctgttct	420
gacaacactg	aagtagaagt	ttctaacttg	gaaaataaac	aaccagttga	gtcgacatct	480
gcaaaatctt	gttctccaag	tcctgtgtct	cctcaggtgc	agccacaagc	agcagatacc	540
atcagtgatt	ctgttgctgt	cccgccatca	ctgctgggca	tgaggagagg	gctgaactca	600
agattggaag	caactgcagc	ctcctcagtt	aaaacacgta	tgcaaaaact	tgcaagagcaa	660
cggcgcggtt	gggataatga	tgatatgaca	gatgacattc	ctgaaagctc	actctttctca	720
ccaatgccat	cagaggaaaa	ggctgcttcc	cctcccagac	ctctgctttc	aaatgcctcg	780
gcaactccag	ttggcagaag	gggcccgtctg	gccaatcttg	ctgcaactat	ttgctcctgg	840
gaagatgatg	taaatcactc	atgttgcaaaa	caaaacagtg	tacaagaaca	gcctgggtacc	900
gcttggtttat	ccaaattttc	ctctgcaagt	ggagcatctg	ctaggatcaa	tagcagcagt	960
gttaagcagg	aagctacatt	ctgttcccaa	agggatggcg	atgcctcttt	gaataaagcc	1020
ctatcctcaa	gtgctgatga	tgcgtctttg	gttaatgcct	caattttccag	ctctgtgaaa	1080
gctactttctc	cagtgaatac	tactacatct	atcactgatg	ctaaaagttg	tgagggacaa	1140
aatcctgagc	tacttccaaa	aactcctatt	agtcctctga	aaacgggggt	atcgaaacca	1200
attgtgaagt	caactttatc	ccagacagtt	ccatccaagg	gagaattaag	tagagaaatt	1260
tgtctgcaat	ctcaatctaa	agacaaatct	acgacaccag	gaggaacagg	aattaagcct	1320
ttcctggaac	gctttggaga	gcgttgtcaa	gaacatagca	aagaaagtcc	agctcgtagc	1380
acaccccaca	gaacccccat	tattactcca	aatacaaagg	ccatccaaga	aagattatc	1440
aagcaagaca	catcttcatc	tactacccat	ttagcacaac	agctcaagca	ggaacgtcaa	1500
aaagaactag	catgtcttcg	tggccgattt	gacaagggca	atatatggag	tcagaaaaaa	1560
ggcggaaact	caaaaagcaa	acaactagaa	accaaacagg	aaactcactg	tcagagcact	1620
cccctcaaaa	aacaccaagg	tgtttcaaaa	actcagtcac	ttccagtaac	agaaaagggtg	1680

```

accgaaaacc agataaccagc caaaaattct agtacagaac ctaaagggtt cactgaatgc 1740
gaaatgacga aatctagccc ttgaaaata acattgtttt tagaagagga caaatcctta 1800
aaagtaacat cagacccaaa ggttgagcag aaaattgaag tgatacgtga aattgagatg 1860
agtgtggatg atgatgatat caatagtctg aaagtaatta atgacctctt cagtgatgtc 1920
ctagaggaag gtgaactaga tatggagaag agccaagagg agatggatca agcattagca 1980
gaaagcagcg aagaacagga agatgcactg aatatctcct caatgtcttt acttgcacca 2040
ttggcacaaa cagttggtgt ggtaagtcca gagagtttag tgtccacacc tagactggaa 2100
ttgaaagaca ccagcagaag tgatgaaagt ccaaaaccag gaaaattcca aagaactcgt 2160
gtccctcgag ctgaatctgg tgatagcctt ggttctgaag atcgtgatct tctttacagc 2220
attgatgcat atagatctca aagattcaaa gaaacagaac gtccatcaat aaagcagggtg 2280
attgttcgga aggaagatgt tacttcaaaa ctggatgaaa aaaataatgc ctttcttgt 2340
caagttaata tcaaacagaa aatgcaggaa ctcaataacg aaataaatat gcaacagaca 2400
gtgatctatc aagctagcca ggctcttaac tgctgtgttg atgaagaaca tggaaaaggg 2460
tccctagaag aagctgaagc agaaagactt cttctaattg caactgggaa gagaacactt 2520
ttgattgatg aattgaataa attgaagaac gaaggacctc agaggaagaa taaggctagt 2580
ccccaaagtg aatttatgcc atccaaagga tcagttactt tgtcagaaat ccgcttgctt 2640
ctaaaagcag attttgtctg cagtacgggt cagaaaccag atgcagcaaa ttactattac 2700
ttaattatac taaaagcagg agctgaaaat atggtagcca caccattagc aagtacttca 2760
aactctctta acggtgatgc tctgacattc actactacat ttactctgca agatgtatcc 2820
aatgactttg aaataaatat tgaagtttac agcttgggtg aaaagaaaga tccctcaggc 2880
cttgataaga agaaaaaaac atccaagtcc aaggctatta ctccaaagcg actcctcaca 2940
tctataacca caaaaagcaa cactcattct tcagtcattg ccagtcagg aggtcttagt 3000
gctgtgcgaa ccagcaactt cgcccttggt ggactctaca cattatcatt gtcttcagta 3060
ggaaatacta agtttgttct ggacaagggt ccccttttat cttctttgga aggtcatatt 3120
tatttaaaaa taaaatgtca agtgaattcc agtgttgaag aaagagggtt tctaaccata 3180
tttgaagatg ttagtggttt tgggtgcctg catcgaagat ggtgtgttct ttctggaaac 3240
tgtatatctt attggactta tccagatgat gagaaacgca agaatcccat aggaaggata 3300
aatctggcta attgtaccag tcgtcagata gaaccagcca acagagaatt ttgtgcaaga 3360
cgcaacactt ttgaattaat tactgtccga ccacaaagag aagatgaccg agagactctt 3420
gtcagccaat gcagggacac actctgtgtt accaagaact ggctgtctgc agatactaaa 3480
gaagagcggg atctctggat gcaaaaactc aatcaagttc ttgttgatat tcgcctctgg 3540
caacctgatg cttgctacaa acctattgga aagccttaaa ccgggaaatt tccatgctat 3600
ctagagggtt ttgatgtcat ctttaagaaac acacttaaga gcatcagatt tactgattgc 3660
attttatgct ttaagtacga aagggtttgt gccaatattc actacgtatt atgcagtatt 3720
tatatctttt gtagtataaa ctttaactga tttctgtcat tcatcaatga gtagaagtaa 3780
atacattata gttgattttg cttaaatctta atttaaaagc ctcatcttcc tagaaatcta 3840
attattcagt tattcatgac aatatttttt taaaagtaag aaattctgag ttgtcttctt 3900
ggagctgtag gtcttgaagc agcaacgtct ttcaggggtt ggagacagaa acccattctc 3960
caatctcagt agttttttcg aaaggctgtg atcatttatt gatcgtgata tgacttggtta 4020
ctaggggtact gaaaaaaatg tctaaggcct ttacagaaac attttttagta atgaggatga 4080
gaactttttc aaatagcaaa tatatattgg cttaaagcat gaggtgtctt tcagaaaagt 4140
gatgtggaca taggaggcaa tgtgtgagac ttgggggttc aatattttat atagaagagt 4200
taataagcac atggtttaca tttaactcagc tactatatat gcagtggtgt gcacattttc 4260
acagaattct ggcttcatta agatcattat ttttgcgtgc tagcttacag acttagcata 4320
ttagtttttt ctactcctac aagtgtaaat tgaaaaatct ttatattaaa aaagtaaaact 4380
gttatgaagc tgctatgtac taataatact ttgcttgcca aagtgtttgg gttttgttgt 4440
tgtttggttg tttggttggt tttggttcat gaacaacagt gtctagaaac ccattttgaa 4500
agtggaaaaa tattaagtca cctatcacct ttaaagcctt ttttttaaaa ttataaaata 4560
ttgtaaagca gggctcacaac ttttaaatat actttgaact tcttctctga attattaaag 4620
ttctttatga cctcatttat aaacactaaa ttctgtcacc tcctgtcatt ttatttttta 4680
ttcattcaaa tgtatttttt cttgtgcata ttataaaaa atatttttat agctcttact 4740
caaataaata cctgtaaatg tctaaaggaa aaaaaaaaaa aaaaaa 4786

```

&lt;210&gt; 465

&lt;211&gt; 76

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 465

```

Met Ser Lys Ala His Pro Pro Glu Leu Lys Lys Phe Met Asp Lys Lys
1           5           10           15

```

Leu Ser Leu Lys Leu Asn Gly Gly Arg His Val Gln Gly Ile Leu Arg  
 20 25 30  
 Gly Phe Asp Pro Phe Met Asn Leu Val Ile Asp Glu Cys Val Glu Met  
 35 40 45  
 Ala Thr Ser Gly Gln Gln Asn Asn Ile Gly Met Val Val Ile Arg Gly  
 50 55 60  
 Asn Ser Ile Ile Met Leu Glu Ala Leu Glu Arg Val  
 65 70 75

&lt;210&gt; 466

&lt;211&gt; 606

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 466

agacgcaaga cgccgggcct acagcgggag cgtgaggaaa gccgtgcgtt gcgttccaag 60  
 gcatctgtga gcccgcgagg tatacaccat gagcaaagct caccctccc agttgaaaaa 120  
 atttatggac aagaagttat cattgaaatt aaatgggtggc agacatgtcc aaggaatatt 180  
 gcggggattt gatcccttta tgaaccttgt gatagatgaa tgtgtggaga tggcgactag 240  
 tggacaacag aacaatattg gaatgggtgt aatacgagga aatagtatca tcatgttaga 300  
 agccttgga cagagtataaa taatggctgt tcagcagaga aacccatgtc ctctctccat 360  
 agggcctgtt ttactatgat gtaaaaatta ggcatgtac attttcatat tagacttttt 420  
 gttaaataaa cttttgtaat agtcaaaaat gctttctcag atgttctgaa tatagaatat 480  
 cagctctcat tccagttttt tctaacatga attttcctgg ttgacattga tttcaaaggg 540  
 ttttatgcat taaagtgaag gaatcttatt aaatgcgaaa aaaaaaaaaa aaaaaaaaaa 600  
 aaaaaa 606

&lt;210&gt; 467

&lt;211&gt; 971

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 467

Met Glu Leu Ser Asp Ala Asn Leu Gln Thr Leu Thr Glu Tyr Leu Lys  
 1 5 10 15  
 Lys Thr Leu Asp Pro Asp Pro Ala Ile Arg Arg Pro Ala Glu Lys Phe  
 20 25 30  
 Leu Glu Ser Val Glu Gly Asn Gln Asn Tyr Pro Leu Leu Leu Thr  
 35 40 45  
 Leu Leu Glu Lys Ser Gln Asp Asn Val Ile Lys Val Cys Ala Ser Val  
 50 55 60  
 Thr Phe Lys Asn Tyr Ile Lys Arg Asn Trp Arg Ile Val Glu Asp Glu  
 65 70 75 80  
 Pro Asn Lys Ile Cys Glu Ala Asp Arg Val Ala Ile Lys Ala Asn Ile  
 85 90 95  
 Val His Leu Met Leu Ser Ser Pro Glu Gln Ile Gln Lys Gln Leu Ser  
 100 105 110  
 Asp Ala Ile Ser Ile Ile Gly Arg Glu Asp Phe Pro Gln Lys Trp Pro  
 115 120 125  
 Asp Leu Leu Thr Glu Met Val Asn Arg Phe Gln Ser Gly Asp Phe His  
 130 135 140  
 Val Ile Asn Gly Val Leu Arg Thr Ala His Ser Leu Phe Lys Arg Tyr  
 145 150 155 160  
 Arg His Glu Phe Lys Ser Asn Glu Leu Trp Thr Glu Ile Lys Leu Val  
 165 170 175  
 Leu Asp Ala Phe Ala Leu Pro Leu Thr Asn Leu Phe Lys Ala Thr Ile  
 180 185 190  
 Glu Leu Cys Ser Thr His Ala Asn Asp Ala Ser Ala Leu Arg Ile Leu  
 195 200 205  
 Phe Ser Ser Leu Ile Leu Ile Ser Lys Leu Phe Tyr Ser Leu Asn Phe

210		215		220
Gln Asp Leu Pro Glu Phe Phe Glu Asp Asn Met Glu Thr Trp Met Asn				
225		230		235
Asn Phe His Thr Leu Thr Leu Asp Asn Lys Leu Leu Gln Thr Asp				240
	245		250	255
Asp Glu Glu Glu Ala Gly Leu Leu Glu Leu Lys Ser Gln Ile Cys				
	260		265	270
Asp Asn Ala Ala Leu Tyr Ala Gln Lys Tyr Asp Glu Glu Phe Gln Arg				
	275		280	285
Tyr Leu Pro Arg Phe Val Thr Ala Ile Trp Asn Leu Leu Val Thr Thr				
	290		295	300
Gly Gln Glu Val Lys Tyr Asp Leu Leu Val Ser Asn Ala Ile Gln Phe				
305		310		315
Leu Ala Ser Val Cys Glu Arg Pro His Tyr Lys Asn Leu Phe Glu Asp				
	325		330	335
Gln Asn Thr Leu Thr Ser Ile Cys Glu Lys Val Ile Val Pro Asn Met				
	340		345	350
Glu Phe Arg Ala Ala Asp Glu Glu Ala Phe Glu Asp Asn Ser Glu Glu				
	355		360	365
Tyr Ile Arg Arg Asp Leu Glu Gly Ser Asp Ile Asp Thr Arg Arg Arg				
	370		375	380
Ala Ala Cys Asp Leu Val Arg Gly Leu Cys Lys Phe Phe Glu Gly Pro				
385		390		395
Val Thr Gly Ile Phe Ser Gly Tyr Val Asn Ser Met Leu Gln Glu Tyr				
	405		410	415
Ala Lys Asn Pro Ser Val Asn Trp Lys His Lys Asp Ala Ala Ile Tyr				
	420		425	430
Leu Val Thr Ser Leu Ala Ser Lys Ala Gln Thr Gln Lys His Gly Ile				
	435		440	445
Thr Gln Ala Asn Glu Leu Val Asn Leu Thr Glu Phe Phe Val Asn His				
	450		455	460
Ile Leu Pro Asp Leu Lys Ser Ala Asn Val Asn Glu Phe Pro Val Leu				
465		470		475
Lys Ala Asp Gly Ile Lys Tyr Ile Met Ile Phe Arg Asn Gln Val Pro				
	485		490	495
Lys Glu His Leu Leu Val Ser Ile Pro Leu Leu Ile Asn His Leu Gln				
	500		505	510
Ala Glu Ser Ile Val Val His Thr Tyr Ala Ala His Ala Leu Glu Arg				
	515		520	525
Leu Phe Thr Met Arg Gly Pro Asn Asn Ala Thr Leu Phe Thr Ala Ala				
	530		535	540
Glu Ile Ala Pro Phe Val Glu Ile Leu Leu Thr Asn Leu Phe Lys Ala				
545		550		555
Leu Thr Leu Pro Gly Ser Ser Glu Asn Glu Tyr Ile Met Lys Ala Ile				
	565		570	575
Met Arg Ser Phe Ser Leu Leu Gln Glu Ala Ile Ile Pro Tyr Ile Pro				
	580		585	590
Thr Leu Ile Thr Gln Leu Thr Gln Lys Leu Leu Ala Val Ser Lys Asn				
	595		600	605
Pro Ser Lys Pro His Phe Asn His Tyr Met Phe Glu Ala Ile Cys Leu				
	610		615	620
Ser Ile Arg Ile Thr Cys Lys Ala Asn Pro Ala Ala Val Val Asn Phe				
625		630		635
Glu Glu Ala Leu Phe Leu Val Phe Thr Glu Ile Leu Gln Asn Asp Val				
	645		650	655
Gln Glu Phe Ile Pro Tyr Val Phe Gln Val Met Ser Leu Leu Leu Glu				
	660		665	670
Thr His Lys Asn Asp Ile Pro Ser Ser Tyr Met Ala Leu Phe Pro His				
	675		680	685
Leu Leu Gln Pro Val Leu Trp Glu Arg Thr Gly Asn Ile Pro Ala Leu				
	690		695	700

Val Arg Leu Leu Gln Ala Phe Leu Glu Arg Gly Ser Asn Thr Ile Ala  
 705 710 715 720  
 Ser Ala Ala Ala Asp Lys Ile Pro Gly Leu Gly Val Phe Gln Lys  
 725 730 735  
 Leu Ile Ala Ser Lys Ala Asn Asp His Gln Gly Phe Tyr Leu Leu Asn  
 740 745 750  
 Ser Ile Ile Glu His Met Pro Pro Glu Ser Val Asp Gln Tyr Arg Lys  
 755 760 765  
 Gln Ile Phe Ile Leu Leu Phe Gln Arg Leu Gln Asn Ser Lys Thr Thr  
 770 775 780  
 Lys Phe Ile Lys Ser Phe Leu Val Phe Ile Asn Leu Tyr Cys Ile Lys  
 785 790 795 800  
 Tyr Gly Ala Leu Ala Leu Gln Glu Ile Phe Asp Gly Ile Gln Pro Lys  
 805 810 815  
 Met Phe Gly Met Val Leu Glu Lys Ile Ile Ile Pro Glu Ile Gln Lys  
 820 825 830  
 Val Ser Gly Asn Val Glu Lys Lys Ile Cys Ala Val Gly Ile Thr Lys  
 835 840 845  
 Leu Leu Thr Glu Cys Pro Pro Met Met Asp Thr Glu Tyr Thr Lys Leu  
 850 855 860  
 Trp Thr Pro Leu Leu Gln Ser Leu Ile Gly Leu Phe Glu Leu Pro Glu  
 865 870 875 880  
 Asp Asp Thr Ile Pro Asp Glu Glu His Phe Ile Asp Ile Glu Asp Thr  
 885 890 895  
 Pro Gly Tyr Gln Thr Ala Phe Ser Gln Leu Ala Phe Ala Gly Lys Lys  
 900 905 910  
 Glu His Asp Pro Val Gly Gln Met Val Asn Asn Pro Lys Ile His Leu  
 915 920 925  
 Ala Gln Ser Leu His Lys Leu Ser Thr Ala Cys Pro Gly Arg Val Pro  
 930 935 940  
 Ser Met Val Ser Thr Ser Leu Asn Ala Glu Ala Leu Gln Tyr Leu Gln  
 945 950 955 960  
 Gly Tyr Leu Gln Ala Ala Ser Val Thr Leu Leu  
 965 970

<210> 468  
 <211> 3579  
 <212> DNA  
 <213> Homo sapiens

<400> 468  
 tcaggctcgc tgtcgcgccca ttttgccggg gtttgaatgt gaggcggagc ggccggcagga 60  
 gcgggtagtg ccagctacgg tccgcgggctg gggttccctc ctccggtttct gtatccccac 120  
 gagatcctat agcaatggaa ctcagcgatg caaatctgca aacactaaca gaatatttaa 180  
 agaaaacact tgatcctgat cctgccatcc gacgtccagc tgagaaattt cttgaatctg 240  
 ttgaaggaaa tcagaattat ccactgttgc ttttgacatt actggagaag tcccaggata 300  
 atgttatcaa agtatgtgct tcagtaacat tcaaaaacta tattaagaag aactggagaa 360  
 ttgttgaaga tgaaccaaac aaaatttgtg aagccgatcg agtggccatt aaagccaaca 420  
 tagtgcactt gatgcttagc agcccagagc aaattcagaa gcagttaagt gatgcaatta 480  
 gcattatttg cagagaagat tttccacaga aatggcctga cttgctgaca gaaatgggtga 540  
 atcgctttca gaggggagat ttccatgtta ttaattggagt cctccgtaca gcacattcat 600  
 tatttaaaag ataccgtcat gaatttaagt caaacgagtt atggactgaa attaagcttg 660  
 ttctggatgc ctttgctttg cctttgacta atctttttta ggccactatt gaactctgca 720  
 gtacccatgc aaatgatgcc tctgccttga ggattctgtt ttcttcctg atcctgatct 780  
 caaaattggt ctatagttaa aactttcagg atctccctga attttttgaa gataatatgg 840  
 aaacttggat gaataatttt catactctct taacattgga taataagctt ttacaaactg 900  
 atgatgaaga ggaagccggc ttattggagc cctttaaata ccagatttgt gataatgccg 960  
 cactctatgc acaaaagtac gatgaagaat tccagcgata cctgcctcgt ttgtttacag 1020  
 ccatctggaa ttactagtt acaacgggtc aagaggttta atatgatttg ttggttaaga 1080  
 atgcaattca atttctggct tcagtttgtg agagacctca ttataagaat ctatttgagg 1140



```

accagaacac gctgacaagt atctgtgaaa aggttattgt gcctaacatg gaatttagag 1200
ctgctgatga agaagcattt gaagataatt ctgaggagta cataaggaga gatttggag 1260
gatctgatat tgatactaga cgcagggctg ctgtgtatct ggtacgagga ttatgcaagt 1320
tttttgaggg acctgtgaca ggaatcttct ctggttatgt taattccatg ctgcaggaat 1380
acgcaaaaaa tccatctgtc aactggaaac acaaagatgc agccatctac ctagtgcacat 1440
ctttggcatc aaaagcccaa acacagaagc atggaattac acaagcaaat gaacttgtaa 1500
acctaactga gttctttgtg aatcacatcc tccctgattt aaaatcagct aatgtgaatg 1560
aatttctgt ccttaaagct gacggtatca aatatattat gatttttaga aatcaagtgc 1620
caaaagaaca tcttttagtc tgcattcttc tcttgattaa tcattctcaa gctgaaagta 1680
ttgttggtca tacttaacga gctcatgctc ttgaacggct ctttactatg cgagggccta 1740
acaatgccac tctctttaca gctgcagaaa tcgcaccgtt tgttgagatt ctgctaacaa 1800
accttttcaa agctctcaca ctctctggct cttcagaaaa tgaatatatt atgaaagcta 1860
tcatgagaag tttttctctc ctacaagaag ccataatccc ctacatccct actctcatca 1920
ctcagcttac acagaagcta ttagctgtta gtaagaaccc aagcaaacct cactttaatc 1980
actacatgtt tgaagaata gttttatcca taagaataac ttgcaaagct aacctgtgc 2040
ctgttgtaaa ttttgaggag gctttgtttt tgggtgtttac tgaaatctta caaatgatg 2100
tgcaagaatt tattccatac gtctttcaag tgatgtcttt gcttctggaa acacacaaaa 2160
atgacatccc gtcttctat atggccttat tctctcatct ccttcagcca gtgctttggg 2220
aaagaacagg aaatattcct gctctagtga ggcttcttca agcattctta gaacgcggtt 2280
caaacacaaat agcaagtgtc gcagctgaca aaattcctgg gttactagggt gtctttcaga 2340
agctgattgc atccaaagca aatgaccacc aagggtttta tcttctaaac agtataatag 2400
agcacatgcc tctgaatca gttgaccaat ataggaaaca aatcttcatt ctgctattcc 2460
agagacttca gaattccaaa acaaccaagt ttatcaagag ttttttagtc tttattaatt 2520
tgtattgcat aaaatatggg gcaactagcac tacaagaat atttgatgg atacaacca 2580
aaatgtttgg aatggttttg gaaaaaatta ttattcctga aattcagaag gtatctggaa 2640
atgtagagaa aaagatctgt gcggttgcca taaccaaat actaacagaa tgtcccccac 2700
tgatggacac tgagtatacc aaactgtgga ctccattatt acagtctttg attgggtctt 2760
ttgagttacc cgaagatgat accattcctg atgaggaaca ttttattgac atagaagata 2820
caccaggata tcagactgcc ttctcacagt tggcatttgc tgggaaaaaa gagcatgatc 2880
ctgtaggtea aatggtgaat aacccccaaa ttcacctggc acagtcactt cacaagttgt 2940
ctaccgcctg tccaggaagg gttccatcaa tgggtgagcac cagcctgaat gcagaagcgc 3000
tccagtatct ccaaggttac cttcaggcag ccagtgtgac actgctttta actgcatttt 3060
tctaattggc taaaccacaga tggtttctta ggaaatcaca ggcttctgag cacagctgca 3120
ttaaaacaaa ggaagttctc cttttgaact tgtcacgaat tccatcttgt aaaggatatt 3180
aaatgttgt ttaacctgaa ccttgagcaa attagttgg ttgtgtgatc atacagttat 3240
gtgggtggct tctagtttgc aacttcaagg gacaagtatt aatagttcag tgtatggcgt 3300
tggtttgtgt tgagcgtttg cacggtttgg ataattctaa attttgacgg acactgtgga 3360
gactttctgt tactaaatcc ttttgttttg aagctgttgc tatttgtatt tctctgttcc 3420
tttatatttt ttgtctgttt atttacgctt ttattggaaa tgtgaataag taaagaatta 3480
cttgtgttac ttgccaagca gtgcacattt catagtttca aatctgtaat cagcaataaa 3540
aatcctaaaa tatgtaccta aaaaaaaaaa aaaaaaaaaa 3579

```

&lt;210&gt; 469

&lt;211&gt; 234

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 469

```

Met Ser Cys Ile Asn Leu Pro Thr Val Leu Pro Gly Ser Pro Ser Lys
 1             5             10             15
Thr Arg Gly Gln Ile Gln Val Ile Leu Gly Pro Met Phe Ser Gly Lys
          20             25             30
Ser Thr Glu Leu Met Arg Arg Val Arg Arg Phe Gln Ile Ala Gln Tyr
          35             40             45
Lys Cys Leu Val Ile Lys Tyr Ala Lys Asp Thr Arg Tyr Ser Ser Ser
          50             55             60
Phe Cys Thr His Asp Arg Asn Thr Met Glu Ala Leu Pro Ala Cys Leu
65             70             75             80
Leu Arg Asp Val Ala Gln Glu Ala Leu Gly Val Ala Val Ile Gly Ile
          85             90             95
Asp Glu Gly Gln Phe Phe Pro Asp Ile Met Glu Phe Cys Glu Ala Met

```

100	105	110
Ala Asn Ala Gly Lys Thr Val Ile	Val Ala Ala Leu Asp Gly Thr Phe	
115	120	125
Gln Arg Lys Pro Phe Gly Ala Ile	Leu Asn Leu Val Pro Leu Ala Glu	
130	135	140
Ser Val Val Lys Leu Thr Ala Val	Cys Met Glu Cys Phe Arg Glu Ala	
145	150	155
Ala Tyr Thr Lys Arg Leu Gly Thr	Glu Lys Glu Val Glu Val Ile Gly	
165	170	175
Gly Ala Asp Lys Tyr His Ser Val	Cys Arg Leu Cys Tyr Phe Lys Lys	
180	185	190
Ala Ser Gly Gln Pro Ala Gly Pro	Asp Asn Lys Glu Asn Cys Pro Val	
195	200	205
Pro Gly Lys Pro Gly Glu Ala Val	Ala Ala Arg Lys Leu Phe Ala Pro	
210	215	220
Gln Gln Ile Leu Gln Cys Ser Pro	Ala Asn	
225	230	

&lt;210&gt; 470

&lt;211&gt; 1421

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 470

acttactgcg	ggacggcctt	ggagagtact	cggttctgtg	aacttcccgg	aggcgcaatg	60
agctgcatta	acctgcccac	tgtgctgccc	ggctccccca	gcaagaccgg	ggggcagatc	120
caggtgatcc	tcgggcccga	gttctcagga	aaaagcacag	agttgatgag	acgcgtccgt	180
cgcttccaga	ttgctcagta	caagtgcctg	gtgatcaagt	atgccaaaga	cactcgtctac	240
agcagcagct	tctgcacaca	tgaccggaac	accatggagg	cgctgcccgc	ctgcctgctc	300
cgagacgtgg	cccaggaggc	cctgggctgt	gctgtcatag	gcacgcagca	ggggcagttt	360
ttccctgaca	tcatggagtt	ctgcgaggcc	atggccaacg	ccgggaagac	cgtaattgtg	420
gctgactggt	atgggacctt	ccagagggaag	ccatttgggg	ccatcctgaa	cctggtgccg	480
ctggccgaga	gcgtgggtgaa	gctgacggcg	gtgtgcatgg	agtgttccg	ggaagccgcc	540
tataccaaga	ggctcggcac	agagaaggag	gtcgaggtga	ttgggggagc	agacaagtac	600
cactccgtgt	gtcggctctg	ctacttcaag	aaggcctcag	gccagcctgc	cgggccggac	660
aacaaagaga	actgcccagt	gccaggaaag	ccagggggaag	ccgtggctgc	caggaagctc	720
tttgcaccac	agcagattct	gcaatgcagc	cctgccaact	gagggacctg	caagggccgc	780
ccgctccctt	cctgcccactg	ccgcctactg	gacgctgccc	tgcatgctgc	ccagccactc	840
caggaggaag	tcgggaggcg	tggagggtga	ccacaccttg	gccttctggg	aactctcctt	900
tgtgtggctg	ccccacctgc	cgcatgctcc	ctcctctcct	accactggt	ctgcttaaa	960
cttccctctc	agctgctggg	acgatcgccc	aggctggagc	tggccccgct	tgggtggcctg	1020
ggatctggca	cactccctct	ccttgggggtg	agggacagag	ccccacgctg	ttgacatcag	1080
cctgcttctt	ccccctgctg	gctttcactg	ctgagtttct	gttctccctg	ggaagcctgt	1140
gccagcacct	ttgagccttg	gcccacactg	aggcttaggc	ctctctgcct	gggatgggct	1200
cccaccctcc	cctgaggatg	gcctggattc	acgcctctt	gtttcctttt	gggctcaaag	1260
cccttccctac	ctctgggtgat	ggtttccaca	ggaacaacag	catctttcac	caagatgggt	1320
ggcaccaacc	ttgctgggac	ttggatccca	ggggcttatc	tcttcaagtg	tggagagggc	1380
agggtccacg	cctctgctgt	agcttatgaa	attaactaat	t		1421

&lt;210&gt; 471

&lt;211&gt; 792

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 471

Met His Val Ile Lys Arg Asp Gly Arg	Gln Glu Arg Val Met Phe Asp
1	5 10 15
Lys Ile Thr Ser Arg Ile Gln Lys Leu	Cys Tyr Gly Leu Asn Met Asp
20	25 30
Phe Val Asp Pro Ala Gln Ile Thr Met	Lys Val Ile Gln Gly Leu Tyr

-230-

Lys Gln Ile Phe Glu Thr Ile Tyr Tyr Gly Ala Leu Glu Ala Ser Cys  
 530 535 540  
 Asp Leu Ala Lys Glu Gln Gly Pro Tyr Glu Thr Tyr Glu Gly Ser Pro  
 545 550 555 560  
 Val Ser Lys Gly Ile Leu Gln Tyr Asp Met Trp Asn Val Thr Pro Thr  
 565 570 575  
 Asp Leu Trp Asp Trp Lys Val Leu Lys Glu Lys Ile Ala Lys Tyr Gly  
 580 585 590  
 Ile Arg Asn Ser Leu Leu Ile Ala Pro Met Pro Thr Ala Ser Thr Ala  
 595 600 605  
 Gln Ile Leu Gly Asn Asn Glu Ser Ile Glu Pro Tyr Thr Ser Asn Ile  
 610 615 620  
 Tyr Thr Arg Arg Val Leu Ser Gly Glu Phe Gln Ile Val Asn Pro His  
 625 630 635 640  
 Leu Leu Lys Asp Leu Thr Glu Arg Gly Leu Trp His Glu Glu Met Lys  
 645 650 655  
 Asn Gln Ile Ile Ala Cys Asn Gly Ser Ile Gln Ser Ile Pro Glu Ile  
 660 665 670  
 Pro Asp Asp Leu Lys Gln Leu Tyr Lys Thr Val Trp Glu Ile Ser Gln  
 675 680 685  
 Lys Thr Val Leu Lys Met Ala Ala Glu Arg Gly Ala Phe Ile Asp Gln  
 690 695 700  
 Ser Gln Ser Leu Asn Ile His Ile Ala Glu Pro Asn Tyr Gly Lys Leu  
 705 710 715 720  
 Thr Ser Met His Phe Tyr Gly Trp Lys Gln Gly Leu Lys Thr Gly Met  
 725 730 735  
 Tyr Tyr Leu Arg Thr Arg Pro Ala Ala Asn Pro Ile Gln Phe Thr Leu  
 740 745 750  
 Asn Lys Glu Lys Leu Lys Asp Lys Glu Lys Val Ser Lys Glu Glu Glu  
 755 760 765  
 Glu Lys Glu Arg Asn Thr Ala Ala Met Val Cys Ser Leu Glu Asn Arg  
 770 775 780  
 Asp Glu Cys Leu Met Cys Gly Ser  
 785 790

<210> 472  
 <211> 3117  
 <212> DNA  
 <213> Homo sapiens

<400> 472  
 attcgaaccc cgtcgcgccc ctttgtgcgt cacgggtggc gggcgcgggg aggggatttg 60  
 gattgttgcg cctctgctct gaagaaagt ctgtctggct ccaactccag ttctttcccc 120  
 tgagcagcgc ctggaacctt acccttccca ctctgtcacc ttctcgatcc cgccggcgct 180  
 ttagagccgc agtccagtct tggatccttc agagcctcag ccactagctg cgatgcatgt 240  
 gatcaagcga gatggccgccc aagaacgagt catgtttgac aaaattacat ctcgaaatcca 300  
 gaagctttgt tatggactca atatggattt tgttgatcct gctcagatca ccatgaaaagt 360  
 aatccaaggc ttgtacagtg gggtcaccac agtggaaacta gatactttgg ctgctgaaac 420  
 agctgcaacc ttgactacta agcaccctga ctatgctatc ctggcagcca ggatcgctgt 480  
 ctctaacttg cacaagaaaa caaagaaagt gttcagtgat gtgatggaag acctctataa 540  
 ctacataaat ccacataatg gcaaacactc tcccatgggtg gccaaagtcaa cattggatat 600  
 tgttctggcc aataaagatc gcctgaattc tgctattatc tatgaccgag atttctctta 660  
 caattacttc ggctttaaga cgctagagcg gtcttatttg ttgaagatca atggaaaagt 720  
 ggctgaaaga ccacaacata tgttgatgag agtatctgtt gggatccaca aagaagacat 780  
 tgatgcagca attgaaacat ataattctct ttctgagagg tggtttactc atgcttcgcc 840  
 cactctcttc aatgctggtt ccaaccgccc acaactttct agctgttttc ttctgagtat 900  
 gaaagatgac agcattggaag gcatttatga cactctaaag caatgtgcat tgatttctaa 960  
 gtctgctgga ggaattgggtg ttgctgtgag ttgtattcgg gctactggca gctacattgc 1020  
 tgggactaat ggcaattcca atggccttgt accgatgctg agagtatata acaacacagc 1080  
 tagatatgtg gatcaagggtg ggaacaagcg tcttggggca tttgctattt acctggagcc 1140

```

ttggcattta gacatctttg aattccttga tttaaagaag aacacaggaa aggaagagca 1200
gcgtgccaga gatcttttct ttgctctttg gattccggat ctcttcatga aacgagtggg 1260
gactaatcag gactgggtctt tgatgtgtcc aaatgagtggt cctgggtctgg atgaggtttg 1320
gggagaggaa tttgagaaac tatatgcaag ttatgagaaa caagggtcgtg tccgcaaagt 1380
tgtaaaagct cagcagctttt ggtatgccat cattgagtgct cagacggaaa caggcaccct 1440
gtatatgctc tacaaagatt cctgtaatcg aaagagcaac cagcagaacc tgggaaccat 1500
caaatgcagc aacctgtgca cagaaatagt ggagtacacc agcaaagatg aggttgctgt 1560
ttgtaatttg gcttccctgg ccctgaatat gtatgtcaca tcagaacaca catacgactt 1620
taagaagttg gctgaagtca ctaaagtcgt tgtccgaaac ttgaataaaa ttattgatat 1680
aaactactat cctgtaccag aggcattgct atcaaataaa cgccatcgcc ccattggaat 1740
tggggtacaa ggtctggcag atgcttttat cctgatgaga tacccttttg agagtgcaga 1800
agcccagtta ctgaataagc agatctttga aactatttat tatggtgctc tggagccag 1860
ctgtgacctt gccaaaggagc agggcccata cgaaacctat gagggctctc cagttagcaa 1920
aggaattctt cagtatgata tgtggaatgt tactctaca gacctatggg actggaaggt 1980
tctcaaggag aagattgcaa agtatggtat aagaaacagt ttacttattg ccccgatgcc 2040
tacagcttcc actgctcaga tcctggggaa taatgagtc attgaacctt acaccagcaa 2100
catctatact cgcagagtct tgtcaggaga atttcagatt gtaaatcctc acttattgaa 2160
agatcttacc gagcggggcc tatggcatga agagatgaaa aaccagatta ttgcatgcaa 2220
tggctctatt cagagcatat cagaaattcc tgatgacctg aagcaacttt ataaaactgt 2280
gtgggaaatc tctcagaaaa ctgttctcaa gatggcagct gagagagggtg ctttcattga 2340
tcaaagccaa tctttgaaca tccacattgc tgagcctaac tatggcaaac tcactagtat 2400
gcacttctac ggctggaagc agggtttgaa gactgggatg tattatttaa ggacaagacc 2460
agcagctaat ccaatccagt tcactctaaa taaggagaag ctaaaagata aagaaaaggt 2520
atcaaaagag gaagaagaga aggagaggaa cacagcagcc atggtgtgct ctttgagaa 2580
tagagatgaa tgtctgatgt gtggatcctg aggaaagact tggagagac cagcatgtct 2640
tcagtagcca aactacttct tgagcataga taggtatagt gggtttgctt gaggtggtaa 2700
ggctttgctg gacctgttg caggcaaaag gtagtaattga tttaaagtac tgttaatgat 2760
gttaatgatt tttttttaa ctcatatatt gggattttca ccaaaataat gcttttgaaa 2820
aaaagaaaaa aaaaacggat atattgagaa tcaaagtaga agtttttaga atgcaaaata 2880
agtcactctt catacagga gtggttaagt aaggtttcat caccattta gcattgcttt 2940
ctgaagactt cagttttgtt aaggagattt agttttactg ctttgactgg tgggtctcta 3000
gaagcaaaac tgagtataa ctcatgagaa gtactgatag gacctttatc tggatatggt 3060
cctataggtt attctgaaat aaagataaac atttctaagt gaaaaaaaaa aaaaaaa 3117

```

<210> 473  
 <211> 674  
 <212> PRT  
 <213> Homo sapiens

<400> 473  
 Met Asn Ser Gly Ala Met Arg Ile His Ser Lys Gly His Phe Gln Gly  
 1 5 10 15  
 Gly Ile Gln Val Lys Asn Glu Lys Asn Arg Pro Ser Leu Lys Ser Leu  
 20 25 30  
 Lys Thr Asp Asn Arg Pro Glu Lys Ser Lys Cys Lys Pro Leu Trp Gly  
 35 40 45  
 Lys Val Phe Tyr Leu Asp Leu Pro Ser Val Thr Ile Ser Glu Lys Leu  
 50 55 60  
 Gln Lys Asp Ile Lys Asp Leu Gly Gly Arg Val Glu Glu Phe Leu Ser  
 65 70 75 80  
 Lys Asp Ile Ser Tyr Leu Ile Ser Asn Lys Lys Glu Ala Lys Phe Ala  
 85 90 95  
 Gln Thr Leu Gly Arg Ile Ser Pro Val Pro Ser Pro Glu Ser Ala Tyr  
 100 105 110  
 Thr Ala Glu Thr Thr Ser Pro His Pro Ser His Asp Gly Ser Ser Phe  
 115 120 125  
 Lys Ser Pro Asp Thr Val Cys Leu Ser Arg Gly Lys Leu Leu Val Glu  
 130 135 140  
 Lys Ala Ile Lys Asp His Asp Phe Ile Pro Ser Asn Ser Ile Leu Ser  
 145 150 155 160  
 Asn Ala Leu Ser Trp Gly Val Lys Ile Leu His Ile Asp Asp Ile Arg

Tyr Tyr Ile Glu Gln Lys Lys Lys Glu Leu Tyr Leu Leu Lys Lys Ser  
 165 170 175  
 180 185 190  
 Ser Thr Ser Val Arg Asp Gly Gly Lys Arg Val Gly Ser Gly Ala Gln  
 195 200 205  
 Lys Thr Arg Thr Gly Arg Leu Lys Lys Pro Phe Val Lys Val Glu Asp  
 210 215 220  
 Met Ser Gln Leu Tyr Arg Pro Phe Tyr Leu Gln Leu Thr Asn Met Pro  
 225 230 235 240  
 Phe Ile Asn Tyr Ser Ile Gln Lys Pro Cys Ser Pro Phe Asp Val Asp  
 245 250 255  
 Lys Pro Ser Ser Met Gln Lys Gln Thr Gln Val Lys Leu Arg Ile Gln  
 260 265 270  
 Thr Asp Gly Asp Lys Tyr Gly Gly Thr Ser Ile Gln Leu Gln Leu Lys  
 275 280 285  
 Glu Lys Lys Lys Lys Gly Tyr Cys Glu Cys Cys Leu Gln Lys Tyr Glu  
 290 295 300  
 Asp Leu Glu Thr His Leu Leu Ser Glu Gln His Arg Asn Phe Ala Gln  
 305 310 315 320  
 Ser Asn Gln Tyr Gln Val Val Asp Asp Ile Val Ser Lys Leu Val Phe  
 325 330 335  
 Asp Phe Val Glu Tyr Glu Lys Asp Thr Pro Lys Lys Lys Arg Ile Lys  
 340 345 350  
 Tyr Ser Val Gly Ser Leu Ser Pro Val Ser Ala Ser Val Leu Lys Lys  
 355 360 365  
 Thr Glu Gln Lys Glu Lys Val Glu Leu Gln His Ile Ser Gln Lys Asp  
 370 375 380  
 Cys Gln Glu Asp Asp Thr Thr Val Lys Glu Gln Asn Phe Leu Tyr Lys  
 385 390 395 400  
 Glu Thr Gln Glu Thr Glu Lys Lys Leu Leu Phe Ile Ser Glu Pro Ile  
 405 410 415  
 Pro His Pro Ser Asn Glu Leu Arg Gly Leu Asn Glu Lys Met Ser Asn  
 420 425 430  
 Lys Cys Ser Met Leu Ser Thr Ala Glu Asp Asp Ile Arg Gln Asn Phe  
 435 440 445  
 Thr Gln Leu Pro Leu His Lys Asn Lys Gln Glu Cys Ile Leu Asp Ile  
 450 455 460  
 Ser Glu His Thr Leu Ser Glu Asn Asp Leu Glu Glu Leu Arg Val Asp  
 465 470 475 480  
 His Tyr Lys Cys Asn Ile Gln Ala Ser Val His Val Ser Asp Phe Ser  
 485 490 495  
 Thr Asp Asn Ser Gly Ser Gln Pro Lys Gln Lys Ser Asp Thr Val Leu  
 500 505 510  
 Phe Pro Ala Lys Asp Leu Lys Glu Lys Asp Leu His Ser Ile Phe Thr  
 515 520 525  
 His Asp Ser Gly Leu Ile Thr Ile Asn Ser Ser Gln Glu His Leu Thr  
 530 535 540  
 Val Gln Ala Lys Ala Pro Phe His Thr Pro Pro Glu Glu Pro Asn Glu  
 545 550 555 560  
 Cys Asp Phe Lys Asn Met Asp Ser Leu Pro Ser Gly Lys Ile His Arg  
 565 570 575  
 Lys Val Lys Ile Ile Leu Gly Arg Asn Arg Lys Glu Asn Leu Glu Pro  
 580 585 590  
 Asn Ala Glu Phe Asp Lys Arg Thr Glu Phe Ile Thr Gln Glu Glu Asn  
 595 600 605  
 Arg Ile Cys Ser Ser Pro Val Gln Ser Leu Leu Asp Leu Phe Gln Thr  
 610 615 620  
 Ser Glu Glu Lys Ser Glu Phe Leu Gly Phe Thr Ser Tyr Thr Glu Lys  
 625 630 635 640  
 Ser Gly Ile Cys Asn Val Leu Asp Ile Trp Glu Glu Glu Asn Ser Asp  
 645 650 655

Asn Leu Leu Thr Ala Phe Phe Ser Ser Pro Ser Thr Ser Thr Phe Thr  
 660 665 670  
 Gly Phe

<210> 474  
 <211> 3729  
 <212> DNA  
 <213> Homo sapiens

<400> 474

```

ggggcacggc gcgtggcgtg agacggggcg gggcgcgcg atcggcgccg cggccgcgtg 60
acgcgttttc aaatcttcaa ccgcgcgcgc cactcgttt gtgctttgcg ccttcctcct 120
ccgcgccttg gagccggatc cggccccgga aaccgcacct gcagacgcgg tacctctact 180
gcgtagaggc cgtagctggc ggaaggagag aggcggccgt cctgtcaaca ggccggggga 240
agccgtgctt tcgcggctgc ccggtgcgac actttctccg gacccagcat gtaggtgccg 300
ggcgactgcc atgaactccg gagccatgag gatccacagt aaaggacatt tccagggtgg 360
aatccaagtc aaaaatgaaa aaaacagacc atctctgaaa tctctgaaaa ctgataacag 420
gccagaaaaa tccaaatgta agccactttg gggaaaagta ttttaccttg acttaccttc 480
tgtcaccata tctgaaaaac ttcaaaagga cattaaggat ctgggagggc gagttgaaga 540
atttctcagc aaagatatca gttatcttat ttcaataaag aaggaagcta aatttgcaca 600
aaccttgggt cgaatttctc ctgtaccaag tccagaatct gcatatactg cagaaaccac 660
ttcacctcat cccagccatg atggaagttc atttaagtca ccagacacag tgtgtttaag 720
cagaggaaaa ttattagttg aaaaagctat caaggacat gattttattc cttcaaatag 780
tatattatca aatgccttgt catggggagt aaaaattcct catattgatg acattagata 840
ctacattgaa caaaagaaaa aagagttgta tttactcaag aaatcaagta cttcagtaag 900
agatgggggc aaaagagttg gtagtggtgc aaaaaaaca agaacaggaa gactcaaaaa 960
gccttttgta aaggtggaag atatgagcca actttatagg ccattttatc ttcagtgac 1020
caatatgcct ttataaaatt attctattca gaagccctgc agtccatttg atgtagacaa 1080
gccatctagt atgcaaaagc aaactcaggt taaactaaga atccaaacag atggcgataa 1140
gtatggtgga acctcaattc aactccagtt gaaagagaag aagaaaaaag gatattgtga 1200
atgttgcttg cagaaatatg aagatctaga aactcacctt ctaagtgagc aacacagaaa 1260
ctttgcacag agtaaccagt atcaagttgt tgatgatatt gtatctaagt tagtttttga 1320
ctttgtggaa tatgaaaagg acacacctaa aaagaaaaga ataaaaatac gtgttggatc 1380
ccttttctct gtttctgcaa gtgtcctgaa aaagactgaa caaaaggaaa aagtggaatt 1440
gcaacatatt tctcagaaag attgccagga agatgataca acagtgaagg agcagaattt 1500
cctgtataaa gagacccagg aaactgaaaa aaagctcctg tttatttcag agcccatccc 1560
ccacccttca aatgaattga gagggcttaa tgagaaaatg agtaataaat gttccatgtt 1620
aagtacagct gaagatgaca taagacagaa ttttacacag ctacctctac ataaaaacaa 1680
acaggaatgc attcttgaca tttccgaaca cacattaagt gaaaatgact tagaagaact 1740
aagggtagat cactataaat gtaacataga ggcattctga catgtttctg atttcagta 1800
agataaatgt ggatctcaac caaaacagaa gtcagatact gtgcttttct cagcaaaagg 1860
tctcaaggaa aaggaccttc attcaatatt tactcatgat tctggtctga taacaataaa 1920
cagttcacaa gagcacctaa ctgttcaggc aaaggctcca ttccatactc ctctgagga 1980
acccaatgaa tgtgacttca agaatatgga tagtttacct tctggtaaaa tacatcgaaa 2040
agtgaaaata atattaggac gaaatagaaa agaaaatctg gaaccaaatt ctgaatttga 2100
taaaagaact gaattttatta cacaagaaga aaacagaatt tgtagttcac cggtagagtc 2160
tttactagac ttgtttcaga ctagtgaaga gaaatcagaa tttttgggtt tcacaagcta 2220
cacagaaaag agtgggtatat gcaatgtttt agatatgttg gaagaggaaa attcagataa 2280
tctgttaaca gcgtttttct cgtcccttc aacttctaca tttactggct tttagaattt 2340
aaaaaatgca tacttttctc agtgataag gatcatattc ttgaaatttt tataaatatg 2400
tatggaaatt cttaggattt ttttaccagc tttgtttaca gacccaaatg taaatattaa 2460
aaataaatat ttgcaatttt ctacagaatt gaatacctgt taaagaaaaa ttacagaata 2520
aacttgtgac tgggtcttgt ttacattata tatattgtcg taattgtttc ctgagaatta 2580
caatgaataa taatttgctt tgtcactgaa aaccaccagt gaagtgaat ttggggaata 2640
tcaaacttag cattatacat ttggatattc tagttgtatt gtaaatttta aaagattatc 2700
aggataacat gaccttgctt tgaaaaacct aaatttcata cagaaaggaa aaatacattt 2760
ctatgcttga attctggaaa tcatcagaaa tattgacact tgagggtcac tgggtggacca 2820
tgctagtatt ttggtgctga gtaattagat tatgtatagg ggtaactgga agccatgaag 2880
gaaggactta accaggctac caagtggaat ataatttatt tgctattttt aattacttac 2940

```

```

cttgtaaata actttaataa tattcaaaag ttgactctcc tgtggctcat gccacaggat 3000
cccagcactt tgggaggccg aggcaggctg atcacgaggt cagaagatca agaccatcct 3060
ggccaacatg gtgaaacctt gtctactaaa aatacaaaaa attagctggg catggcggca 3120
tgcgcctgta gtcccagcta ctccctccag aggctgagaa ttgcttgaac ctgggaagtg 3180
caggttgcag tgagccaaga ttgtgccact gtattccaga ctgggtgaca gagtgaagact 3240
ccatctcaaa aaaaaaaaaa ttgactctac cccctaattt ggtaggagat gaaggagaaa 3300
aggatggcat tgaattatag atacagtttt gggatatata caaggattgc ctgatctgg 3360
cacttactga tacaagcatt tggagaagag aaaattcaaa tataagaaat ttcaatttga 3420
atctgtggat ataaggcatt tgcctttgag tatgttcttg taggagtaag actaccacca 3480
gtgacattca attggttcta ttgcatagta catgaatcta ttgtcactaa aaattagggtt 3540
actaatctta atttttttga aatatatgat tctcagtaca ggactaattc atatgctttt 3600
cctgcataca ttattttgat catttggata acatcaatga acactggctt tttaacacct 3660
ttgtttttca taccagtacc tgaagtaggc tcaataaaaag agtcctttaca gtgaaaaaaa 3720
aaaaaaaaa                                     3729

```

&lt;210&gt; 475

&lt;211&gt; 255

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 475

```

Met Val Lys Leu Thr Ala Glu Leu Ile Glu Gln Ala Ala Gln Tyr Thr
1          5          10          15
Asn Ala Val Arg Asp Arg Glu Leu Asp Leu Arg Gly Tyr Lys Ile Pro
20        25        30
Val Ile Glu Asn Leu Gly Ala Thr Leu Asp Gln Phe Asp Ala Ile Asp
35        40        45
Phe Ser Asp Asn Glu Ile Arg Lys Leu Asp Gly Phe Pro Leu Leu Arg
50        55        60
Arg Leu Lys Thr Leu Leu Val Asn Asn Asn Arg Ile Cys Arg Ile Gly
65        70        75        80
Glu Gly Leu Asp Gln Ala Leu Pro Cys Leu Thr Glu Leu Ile Leu Thr
85        90        95
Asn Asn Ser Leu Val Glu Leu Gly Asp Leu Asp Pro Leu Ala Ser Leu
100       105       110
Lys Ser Leu Thr Tyr Leu Ser Ile Leu Arg Asn Pro Val Thr Asn Lys
115       120       125
Lys His Tyr Arg Leu Tyr Val Ile Tyr Lys Val Pro Gln Val Arg Val
130       135       140
Leu Asp Phe Gln Lys Val Lys Leu Lys Glu Arg Gln Glu Ala Glu Lys
145       150       155       160
Met Phe Lys Gly Lys Arg Gly Ala Gln Leu Ala Lys Asp Ile Ala Arg
165       170       175
Arg Ser Lys Thr Phe Asn Pro Gly Ala Gly Leu Pro Thr Asp Lys Lys
180       185       190
Arg Gly Gly Pro Ser Pro Gly Asp Val Glu Ala Ile Lys Asn Ala Ile
195       200       205
Ala Asn Ala Ser Thr Leu Ala Glu Val Glu Arg Leu Lys Gly Leu Leu
210       215       220
Gln Ser Gly Gln Ile Pro Gly Arg Glu Arg Arg Ser Gly Pro Thr Asp
225       230       235       240
Asp Gly Glu Glu Glu Met Glu Glu Asp Thr Val Thr Asn Gly Ser
245       250       255

```

&lt;210&gt; 476

&lt;211&gt; 1054

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 476



```

gaattccgcg ggaggccacg ggctttccac agcgcgggggg aacggggaggc tgcaggatgg 60
tcaagctgac ggcgagctg atcgagcagg cggcgagta caccaacgcg gtgcgcgacc 120
gggagctgga cctccggggg tataaaattc ccgtcattga aaatctaggt gctacgttag 180
accagtttga tgctattgat ttttctgaca atgagatcag gaaactggat ggttttcctt 240
tgttgagaag actgaaaaca ttgttagtga acaacaacag aatatgccgt ataggtgagg 300
gacttgatca ggctctgccc tgtctgacag aactcattct caccaataat agtctcgtgg 360
aactgggtga tctggaccct ctggcatctc tcaaatcgct gacttaccta agtatcctaa 420
gaaatccggt aaccaataag aagcattaca gattgtatgt gatttataaa gttccgcaag 480
tcagagtact ggatttccag aaagtgaac taaaagagcg tcagggaagca gagaaaatgt 540
tcaagggcaa acgggggtgca cagcttgcaa aggatattgc caggagaagc aaaactttta 600
atccaggtgc tggtttgcca actgacaaaa agagaggtgg gccatctcca ggggatgtag 660
aagcaatcaa gaatgccata gcaaatgctt caactctggc tgaagtggag aggctgaagg 720
ggttgctgca gtctggtcag atccctggca gagaacgcag atcagggccc actgatgatg 780
gtgaagaaga gatggaagaa gacacagtca caaacgggtc ctgagcagtg aggcagatgt 840
ataataatag gccctcttgg aacaagtctt gcttttcgaa catggtataa tagccttgtt 900
tgtgttagca aagtggaatc tctcagcatt gttgaaatgc ttaagactgc tgctgataat 960
tttgtaatat aagttttgaa atctaaatgt caattttcta caaattataa aaataaaactc 1020
cactctctat gctaaaaaaa aaaaaaagga attc 1054

```

&lt;210&gt; 477

&lt;211&gt; 241

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 477

```

Met Arg His Gln Gln Thr Glu Arg Gln Asp Pro Ser Gln Pro Leu Ser
1          5          10          15
Arg Gln His Gly Thr Tyr Arg Gln Ile Phe His Pro Glu Gln Leu Ile
20          25          30
Thr Gly Lys Glu Asp Ala Ala Asn Asn Tyr Ala Trp Gly His Tyr Thr
35          40          45
Ile Gly Lys Glu Phe Ile Asp Leu Leu Leu Asp Arg Ile Arg Lys Leu
50          55          60
Ala Asp Gln Cys Thr Gly Leu Gln Gly Phe Leu Val Phe His Ser Leu
65          70          75          80
Gly Arg Gly Thr Gly Ser Asp Val Thr Ser Phe Leu Met Glu Trp Leu
85          90          95
Ser Val Asn Tyr Gly Lys Lys Ser Lys Leu Gly Phe Ser Ile Tyr Pro
100         105         110
Ala Pro Gln Val Ser Thr Ala Met Val Gln Pro Tyr Asn Ser Ile Leu
115         120         125
Thr Thr His Thr Thr Leu Glu His Ser Asp Cys Ala Phe Met Val Asp
130         135         140
Asn Lys Ala Ile Tyr Asp Ile Cys His Arg Asn Leu Asp Ile Glu Arg
145         150         155         160
Pro Thr Tyr Thr Asn Leu Asn Arg Leu Ile Ser Gln Ile Val Ser Ser
165         170         175
Ile Thr Ala Ser Leu Arg Phe Asp Gly Ala Leu Asn Val Asp Leu Thr
180         185         190
Glu Phe Gln Thr Asn Leu Val Ser Tyr Leu Thr Ser Thr Ser Pro Trp
195         200         205
Pro Pro Met His Gln Ser Ser Leu Gln Lys Lys Tyr Thr Thr Ser Ser
210         215         220
Cys Arg Trp Gln Arg Leu Pro Met Pro Ala Leu Ser Leu Pro Thr Arg
225         230         235         240
Trp

```

&lt;210&gt; 478

&lt;211&gt; 1380

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 478

```

gtctctgccc atccgcgcac cggggcttcg gctggagagg gccagctcgc ttcaggaggc 60
cgaaccccg tccaccaaac cctctcagct cagacgcggg gtgctgagtc acgggggggg 120
ggtggttctg tggatagttg gaatgcatac acagaggaaa gggggatgcy gcaccagcag 180
acagagagac aagaccccgag ccagcccctg tccaggcagc atggcacata ccgccagatc 240
ttccatccag agcagctcat cacaggcaag gaagatgctg ccaataacta tgcctggggc 300
cactacacca ttgggaagga gttcatcgac ctgctactgg accggattcg gaagctggct 360
gaccagtgcg caggacttca gggcttcctg gtgttcacac gccttggtcg gggcactggc 420
tctgacgtca cctcattcct gatggagtgg ctttctgtta actatggcaa gaaatccaag 480
ctgggattct ccatctaccc agccccccag gtgtctacag ccatgggtcca gccctacaac 540
tctatcctga ccaccacac caccctggag cactcagact gtgccttcat ggtggacaac 600
aaagcaatct atgacatctg ccaccgcaac ctagacattg agcgcccaac ctacaccaac 660
ctcaatcgcc tcattagcca aattgtctcc tccatcacag cttctctgcy ctttgacggg 720
gccctcaatg tggacctgac agagttccag accaacctgg tgtcctacct cacatccact 780
tccccctggc cacctatgca ccagtcactc ctgcagaaaa agtataccac gagcagctgt 840
cggtggcaga gattaccaat gcctgctttg agcctgccaa ccagatgggt aagtgtgatc 900
cccggcacgg caagtacatg gcctgctgcc tgcataacca tggagatgtg gtgcccagg 960
atgtcaacgc tgccattgct gccatcaaga ccaagtgcag cattcagttt gtggactggt 1020
gccccacagg cttaaggtt gatatcaatc accagcctcc cactgtgggt cctgggagt 1080
acctggtaaa gtgcaacgtg ccatgtgcat gctgagcaac atgacagcca tcaactatgg 1140
ctgggcccgc ctggaccaca agtttgacct gatgtatgcc aagagggcgt ttgggcaact 1200
atatgtgggt gagggcatgg aggaggggtg gttctccaag gccatgagg atatgactgc 1260
cctggagaag gattacaagg aggtgggcat ggatagtgtg gagtgtgggg aagaaaagat 1320
aggggggatg aatactaggg gaatactgtg tgtctgtcct acataaagtg ctgtggcctt 1380

```

&lt;210&gt; 479

&lt;211&gt; 175

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 479

```

Met Thr Asp Cys Glu Phe Gly Tyr Ile Tyr Arg Leu Ala Gln Asp Tyr
1      5      10      15
Leu Gln Cys Val Leu Gln Ile Pro Gln Pro Gly Ser Gly Pro Ser Lys
20     25     30
Thr Ser Arg Val Leu Gln Asn Val Ala Phe Ser Val Gln Lys Glu Val
35     40     45
Glu Lys Asn Leu Lys Ser Cys Leu Asp Asn Val Asn Val Val Ser Val
50     55     60
Asp Thr Ala Arg Thr Leu Phe Asn Gln Val Met Glu Lys Glu Phe Glu
65     70     75     80
Asp Gly Ile Ile Asn Trp Gly Arg Ile Val Thr Ile Phe Ala Phe Glu
85     90     95
Gly Ile Leu Ile Lys Lys Leu Leu Arg Gln Gln Ile Ala Pro Asp Val
100    105    110
Asp Thr Tyr Lys Glu Ile Ser Tyr Phe Val Ala Glu Phe Ile Met Asn
115    120    125
Asn Thr Gly Glu Trp Ile Arg Gln Asn Gly Gly Trp Glu Asn Gly Phe
130    135    140
Val Lys Lys Phe Glu Pro Lys Ser Gly Trp Met Thr Phe Leu Glu Val
145    150    155    160
Thr Gly Lys Ile Cys Glu Met Leu Ser Leu Leu Lys Gln Tyr Cys
165    170    175

```

&lt;210&gt; 480

&lt;211&gt; 885

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 480

```

agcctacgca cgaaagtgc tagggaggaa ggatattata aagtgatgca aacagaaatt 60
ccaccagcct ccatgtatca tcatgtgtca taactcagtc aagctcagtg agcatttctca 120
gcacattgcc tcaacagctt caaggtgagc cagctcaaga ctttgctctc caccaggcag 180
aagatgacag actgtgaatt tggatatatt tacaggctgg ctcaggacta tctgcagtg 240
gtcctacaga taccacaacc tggatcaggt ccaagcaaaa cgtccagagt gctacaaaat 300
gttgctgtct cagtccaaaa agaagtggaa aagaatctga agtcatgctt ggacaatgtt 360
aatgttgtgt ccgtagacac tggcagaaca ctattcaacc aagtgatgga aaaggagttt 420
gaagacggca tcattaactg gggaagaatt gtaaccatat ttgcatttga aggtattctc 480
atcaagaaac ttctacgaca gcaaattgcc ccggatgtgg atacctataa ggagatttca 540
tattttgttg cggagttcat aatgaataac acaggagaat ggataaggca aaacggaggc 600
tgggaaaatg gctttgtaaa gaagtgtgaa cctaaatctg gctggatgac ttttctagaa 660
gttacaggaa agatctgtga aatgctatct ctcctgaagc aatactgttg accagaaagg 720
acactccata ttgtgaaacc ggcctaattt ttctgactga tatggaaacg attgccaaac 780
catacttcta cttttaataa aacaactttg atgatgtaac ttgaccttcc agagttatgg 840
aaattttgtc cccatgtaat gaataaattg tatgtatttt tctct 885

```

&lt;210&gt; 481

&lt;211&gt; 104

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 481

```

Met Ser His Thr Ile Leu Leu Val Gln Pro Thr Lys Arg Pro Glu Gly
1      5      10      15
Arg Thr Tyr Ala Asp Tyr Glu Ser Val Asn Glu Cys Met Glu Gly Val
20     25     30
Cys Lys Met Tyr Glu Glu His Leu Lys Arg Met Asn Pro Asn Ser Pro
35     40     45
Ser Ile Thr Tyr Asp Ile Ser Gln Leu Phe Asp Phe Ile Asp Asp Leu
50     55     60
Ala Asp Leu Ser Cys Leu Val Tyr Arg Ala Asp Thr Gln Thr Tyr Gln
65     70     75     80
Pro Tyr Asn Lys Asp Trp Ile Lys Glu Lys Ile Tyr Val Leu Leu Arg
85     90     95
Arg Gln Ala Gln Gln Ala Gly Lys
100

```

&lt;210&gt; 482

&lt;211&gt; 815

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 482

```

gcggcggttg agttaagctc gtgtaacggc ggcggtgtcg gcagctgctg tagcgaagag 60
agtttggcgc gatgtctcac accattttgc tggtagagcc taccaagagg ccagaaggca 120
gaacttatgc tgactacgaa tctgtgaatg aatgcatgga aggtgtttgt aaaatgtatg 180
aagaacatct gaaaagaatg aatcccaaca gtccctctat cacatatgac atcagtcagt 240
tgtttgattt catcgatgat ctggcagacc tcagctgcct ggtttaccga gctgataccc 300
agacatacca gccttataac aaagactgga ttaaagagaa gatctacgtg ctcccttcgtc 360
ggcaggccca acaggctggg aaataattgt gttggaagca ctgggggggt tgggggtgggc 420
ttggaacaca ggtgtgtaca gcgtgctgta gtggaagttt tgtatcatag taatcctgtt 480
tccactttgt tatactctag ccaagattga ctgtattaga tgaaatgtga ggatcctgtt 540
caatcggaac ccccggttac ctccctcttt tctttctctt tctttttttt ttttttactt 600
aaacattttt atgatgattt agatggaagt tgttcttcgt cacttaatgt tgggtccagt 660
ccttcaactg ttcatatcta ctttataaca ttcacatact aacccttctt caagatgggg 720
tgggggggtg aaatgcagtt tagccatgtc ctcaagataa agtcttggtg aaaataaata 780

```

aatgtccttt agttataaaa aaaaaaaaaa aaaaa

815

&lt;210&gt; 483

&lt;211&gt; 857

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 483

Met Glu Ser Glu Asp Leu Ser Gly Arg Glu Leu Thr Ile Asp Ser Ile  
 1 5 10 15  
 Met Asn Lys Val Arg Asp Ile Lys Asn Lys Phe Lys Asn Glu Asp Leu  
 20 25 30  
 Thr Asp Glu Leu Ser Leu Asn Lys Ile Ser Ala Asp Thr Thr Asp Asn  
 35 40 45  
 Ser Gly Thr Val Asn Gln Ile Met Met Met Ala Asn Asn Pro Glu Asp  
 50 55 60  
 Trp Leu Ser Leu Leu Leu Lys Leu Glu Lys Asn Ser Val Pro Leu Ser  
 65 70 75 80  
 Asp Ala Leu Leu Asn Lys Leu Ile Gly Arg Tyr Ser Gln Ala Ile Glu  
 85 90 95  
 Ala Leu Pro Pro Asp Lys Tyr Gly Gln Asn Glu Ser Phe Ala Arg Ile  
 100 105 110  
 Gln Val Arg Phe Ala Glu Leu Lys Ala Ile Gln Glu Pro Asp Asp Ala  
 115 120 125  
 Arg Asp Tyr Phe Gln Met Ala Arg Ala Asn Cys Lys Lys Phe Ala Phe  
 130 135 140  
 Val His Ile Ser Phe Ala Gln Phe Glu Leu Ser Gln Gly Asn Val Lys  
 145 150 155 160  
 Lys Ser Lys Gln Leu Leu Gln Lys Ala Val Glu Arg Gly Ala Val Pro  
 165 170 175  
 Leu Glu Met Leu Glu Ile Ala Leu Arg Asn Leu Asn Leu Gln Lys Lys  
 180 185 190  
 Gln Leu Leu Ser Glu Glu Glu Lys Lys Asn Leu Ser Ala Ser Thr Val  
 195 200 205  
 Leu Thr Ala Gln Glu Ser Phe Ser Gly Ser Leu Gly His Leu Gln Asn  
 210 215 220  
 Arg Asn Asn Ser Cys Asp Ser Arg Gly Gln Thr Thr Lys Ala Arg Phe  
 225 230 235 240  
 Leu Tyr Gly Glu Asn Met Pro Pro Gln Asp Ala Glu Ile Gly Tyr Arg  
 245 250 255  
 Asn Ser Leu Arg Gln Thr Asn Lys Thr Lys Gln Ser Cys Pro Phe Gly  
 260 265 270  
 Arg Val Pro Val Asn Leu Leu Asn Ser Pro Asp Cys Asp Val Lys Thr  
 275 280 285  
 Asp Asp Ser Val Val Pro Cys Phe Met Lys Arg Gln Thr Ser Arg Ser  
 290 295 300  
 Glu Cys Arg Asp Leu Val Val Pro Gly Ser Lys Pro Ser Gly Asn Asp  
 305 310 315 320  
 Ser Cys Glu Leu Arg Asn Leu Lys Ser Val Gln Asn Ser His Phe Lys  
 325 330 335  
 Glu Pro Leu Val Ser Asp Glu Lys Ser Ser Glu Leu Ile Ile Thr Asp  
 340 345 350  
 Ser Ile Thr Leu Lys Asn Lys Thr Glu Ser Ser Leu Leu Ala Lys Leu  
 355 360 365  
 Glu Glu Thr Lys Glu Tyr Gln Glu Pro Glu Val Pro Glu Ser Asn Gln  
 370 375 380  
 Lys Gln Trp Gln Ser Lys Arg Lys Ser Glu Cys Ile Asn Gln Asn Pro  
 385 390 395 400  
 Ala Ala Ser Ser Asn His Trp Gln Ile Pro Glu Leu Ala Arg Lys Val  
 405 410 415  
 Asn Thr Glu Gln Lys His Thr Thr Phe Glu Gln Pro Val Phe Ser Val

420					425					430					
Ser	Lys	Gln	Ser	Pro	Pro	Ile	Ser	Thr	Ser	Lys	Trp	Phe	Asp	Pro	Lys
435					440					445					
Ser	Ile	Cys	Lys	Thr	Pro	Ser	Ser	Asn	Thr	Leu	Asp	Asp	Tyr	Met	Ser
450					455					460					
Cys	Phe	Arg	Thr	Pro	Val	Val	Lys	Asn	Asp	Phe	Pro	Pro	Ala	Cys	Gln
465					470					475					
Leu	Ser	Thr	Pro	Tyr	Gly	Gln	Pro	Ala	Cys	Phe	Gln	Gln	Gln	Gln	His
485					490					495					
Gln	Ile	Leu	Ala	Thr	Pro	Leu	Gln	Asn	Leu	Gln	Val	Leu	Ala	Ser	Ser
500					505					510					
Ser	Ala	Asn	Glu	Cys	Ile	Ser	Val	Lys	Gly	Arg	Ile	Tyr	Ser	Ile	Leu
515					520					525					
Lys	Gln	Ile	Gly	Ser	Gly	Gly	Ser	Ser	Lys	Val	Phe	Gln	Val	Leu	Asn
530					535					540					
Glu	Lys	Lys	Gln	Ile	Tyr	Ala	Ile	Lys	Tyr	Val	Asn	Leu	Glu	Glu	Ala
545					550					555					
Asp	Asn	Gln	Thr	Leu	Asp	Ser	Tyr	Arg	Asn	Glu	Ile	Ala	Tyr	Leu	Asn
565					570					575					
Lys	Leu	Gln	Gln	His	Ser	Asp	Lys	Ile	Ile	Arg	Leu	Tyr	Asp	Tyr	Glu
580					585					590					
Ile	Thr	Asp	Gln	Tyr	Ile	Tyr	Met	Val	Met	Glu	Cys	Gly	Asn	Ile	Asp
595					600					605					
Leu	Asn	Ser	Trp	Leu	Lys	Lys	Lys	Lys	Ser	Ile	Asp	Pro	Trp	Glu	Arg
610					615					620					
Lys	Ser	Tyr	Trp	Lys	Asn	Met	Leu	Glu	Ala	Val	His	Thr	Ile	His	Gln
625					630					635					
His	Gly	Ile	Val	His	Ser	Asp	Leu	Lys	Pro	Ala	Asn	Phe	Leu	Ile	Val
645					650					655					
Asp	Gly	Met	Leu	Lys	Leu	Ile	Asp	Phe	Gly	Ile	Ala	Asn	Gln	Met	Gln
660					665					670					
Pro	Asp	Thr	Thr	Ser	Val	Val	Lys	Asp	Ser	Gln	Val	Gly	Thr	Val	Asn
675					680					685					
Tyr	Met	Pro	Pro	Glu	Ala	Ile	Lys	Asp	Met	Ser	Ser	Arg	Glu	Asn	
690					695					700					
Gly	Lys	Ser	Lys	Ser	Lys	Ile	Ser	Pro	Lys	Ser	Asp	Val	Trp	Ser	Leu
705					710					715					
Gly	Cys	Ile	Leu	Tyr	Tyr	Met	Thr	Tyr	Gly	Lys	Thr	Pro	Phe	Gln	Gln
725					730					735					
Ile	Ile	Asn	Gln	Ile	Ser	Lys	Leu	His	Ala	Ile	Ile	Asp	Pro	Asn	His
740					745					750					
Glu	Ile	Glu	Phe	Pro	Asp	Ile	Pro	Glu	Lys	Asp	Leu	Gln	Asp	Val	Leu
755					760					765					
Lys	Cys	Cys	Leu	Lys	Arg	Asp	Pro	Lys	Gln	Arg	Ile	Ser	Ile	Pro	Glu
770					775					780					
Leu	Leu	Ala	His	Pro	Tyr	Val	Gln	Ile	Gln	Thr	His	Pro	Val	Asn	Gln
785					790					795					
Met	Ala	Lys	Gly	Thr	Glu	Glu	Met	Lys	Tyr	Val	Leu	Gly	Gln	Leu	
805					810					815					
Val	Gly	Leu	Asn	Ser	Pro	Asn	Ser	Ile	Leu	Lys	Ala	Ala	Lys	Thr	Leu
820					825					830					
Tyr	Glu	His	Tyr	Ser	Gly	Gly	Glu	Ser	His	Asn	Ser	Ser	Ser	Ser	Lys
835					840					845					
Thr	Phe	Glu	Lys	Lys	Arg	Gly	Lys	Lys							
850					855										

<210> 484  
 <211> 2984  
 <212> DNA  
 <213> Homo sapiens

&lt;400&gt; 484

```

ggaaattcaa acgtgtttgc ggaaaggagt ttgggttcca tcttttcatt tccccagcgc 60
agctttctgt agaaatggaa tccgaggatt taagtggcag agaattgaca attgattcca 120
taatgaacaa agtgagagac attaaaaata agtttaaaaa tgaagacctt actgatgaac 180
taagcttgaa taaaatttct gctgatacta cagataactc gggaactgtt aaccaaatta 240
tgatgatggc aaacaaccca gaggactggg tgagtttgtt gctcaaacta gagaaaaaca 300
gtgttccgct aagtgatgct cttttaaaata aattgattgg tcgttacagt caagcaattg 360
aagcgcttcc cccagataaa tatggccaaa atgagagttt tgctagaatt caagtgaat 420
ttgctgaatt aaaagctatt caagagccag atgatgcacg tgactacttt caaatggcca 480
gagcaaaactg caagaaattt gcttttgttc atatatcttt tgcacaattt gaactgtcac 540
aaggtaatgt caaaaaaagt aaacaacttc ttcaaaaagc tgtagaacgt ggagcagtac 600
cactagaaat gctggaaatt gccctgcgga atttaaacct ccaaaaaaag cagctgcttt 660
cagaggagga aaagaagaat ttatcagcat ctacgggtatt aactgcccga gaatcatttt 720
ccggttcact tgggcattta cagaatagga acaacagttg tgattccaga ggacagacta 780
ctaaagccag gtttttatat ggagagaaca tgccaccaca agatgcagaa ataggttacc 840
ggaaattcatt gagacaaact aacaaaacta aacagtcattg cccatttgga agagtcccag 900
ttaaccttct aaatagccca gattgtgatg tgaagacaga tgattcagtt gtaccttgtt 960
ttatgaaaag acaaacctct agatcagaat gccgagattt ggttgtgcct ggatctaaac 1020
caagtggaaa tgattcctgt gaattaagaa atttaagtc tgttcaaaat agtcatttca 1080
aggaacctct ggtgtcagat gaaaagagtt ctgaacttat tattactgat tcaataaccc 1140
tgaagaataa aacggaatca agtcttctag ctaaattaga agaaactaaa gagtatcaag 1200
aaccagaggt tccagagagt aaccagaaac agtggcaatc taagagaaag tcagagtgt 1260
ttaaccagaa tcctgctgca tcttcaaatc actggcagat tccggagtta gcccgaaaag 1320
ttaatacaga gcagaaacat accacttttg agcaacctgt cttttcagtt tcaaaacagt 1380
caccaccaat atcaacatct aaatggtttg acccaaaatc tatttgtaag acaccaagca 1440
gcaatacctt ggatgattac atgagctgtt ttagaactcc agttgtaag aatgactttc 1500
cacctgcttg tcagttgtca acaccttatg gccaacctgc ctgtttccag cagcaacagc 1560
atcaaatact tgccactcca cttcaaaatt tacaggtttt agcatcttct tcagcaaatg 1620
aatgcatttc ggtaaagga agaatttatt ccattttaaa gcagatagga agtggagggt 1680
caagcaaggt atttcaggtg ttaaattgaa agaaacagat atatgctata aaatatgtga 1740
acttagaaga agcagataac caaactcttg atagttaccg gaacgaaata gcttatttga 1800
ataaactaca acaacacagt gataagatca tccgacttta tgattatgaa atcacggacc 1860
agtacatcta catggtaatg gagtgtggaa atattgatct taatagttgg cttaaaaaga 1920
aaaaatccat tgatccatgg caacgcaaga gtacttgaa aaatatgtta gaggcagttc 1980
acacaatcca tcaacatggc attgttcaca gtgatcttaa accagctaac tttctgatag 2040
ttgatggaaat gctaaagcta attgattttg ggattgcaaa ccaaatgcaa ccagatacaa 2100
caagtgttgt taaagattct caggttggca cagtttaatta tatgccacca gaagcaatca 2160
aagatatgtc ttctccaga gagaatggga aatctaagtc aaagataagc ccaaaaagt 2220
atgtttgtgc cttaggatgt attttgtact atatgactta cgggaaaaca ccatttcagc 2280
agataattaa tcagatttct aaattacatg ccataattga tcctaatacat gaaattgaat 2340
ttcccgatat tccagagaaa gatcttcaag atgtgtttaa gtgttggtta aaaagggacc 2400
caaaacagag gatattccatt cctgagctcc tggtcatcc ctatgttcaa attcaaactc 2460
atccagttaa ccaaatggcc aagggaacca ctgaagaaat gaaatatgtt ctgggccaac 2520
ttgttggctc gaattctcct aactccattt tgaagctgc taaaacttta tatgaacact 2580
atagtgggtg tgaaagtcac aattcttcat cctccaagac ttttgaaaaa aaaaggggaa 2640
aaaaatgatt tgcagttatt cgtaatgtca aataccacct ataaaatata ttggactgtt 2700
atactcttga atccctgtgg aaatctacat ttgaagacaa catcactctg aagtgttatc 2760
agcaaaaaaa attcagtaga ttatctttta aagaaaactg taaaaatagc aaccacttat 2820
ggtactgtat atattgtaga cttgttttct ctgttttatg ctcttggtga atctacttga 2880
catcatttta ctcttggaa agtgggtgga tagcaagtat attctaaaaa actttgttaa 2940
taaagttttg tggctaaaat gacactaaaa aaaaaaaaaa aaaa 2984

```

&lt;210&gt; 485

&lt;211&gt; 725

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 485

```

Met Ala Met Asp Ser Ser Leu Gln Ala Arg Leu Phe Pro Gly Leu Ala
  1                      5                      10                      15

```

Ile Lys Ile Gln Arg Ser Asn Gly Leu Ile His Ser Ala Asn Val Arg  
 20 25 30  
 Thr Val Asn Leu Glu Lys Ser Cys Val Ser Val Glu Trp Ala Glu Gly  
 35 40 45  
 Gly Ala Thr Lys Gly Lys Glu Ile Asp Phe Asp Asp Val Ala Ala Ile  
 50 55 60  
 Asn Pro Glu Leu Leu Gln Leu Leu Pro Leu His Pro Lys Asp Asn Leu  
 65 70 75 80  
 Pro Leu Gln Glu Asn Val Thr Ile Gln Lys Gln Lys Arg Arg Ser Val  
 85 90 95  
 Asn Ser Lys Ile Pro Ala Pro Lys Glu Ser Leu Arg Ser Arg Ser Thr  
 100 105 110  
 Arg Met Ser Thr Val Ser Glu Leu Arg Ile Thr Ala Gln Glu Asn Asp  
 115 120 125  
 Met Glu Val Glu Leu Pro Ala Ala Asn Ser Arg Lys Gln Phe Ser  
 130 135 140  
 Val Pro Pro Ala Pro Thr Arg Pro Ser Cys Pro Ala Val Ala Glu Ile  
 145 150 155 160  
 Pro Leu Arg Met Val Ser Glu Glu Met Glu Glu Gln Val His Ser Ile  
 165 170 175  
 Arg Gly Ser Ser Ala Asn Pro Val Asn Ser Val Arg Arg Lys Ser  
 180 185 190  
 Cys Leu Val Lys Glu Val Glu Lys Met Lys Asn Lys Arg Glu Glu Lys  
 195 200 205  
 Lys Ala Gln Asn Ser Glu Met Arg Met Lys Arg Ala Gln Glu Tyr Asp  
 210 215 220  
 Ser Ser Phe Pro Asn Trp Glu Phe Ala Arg Met Ile Lys Glu Phe Arg  
 225 230 235 240  
 Ala Thr Leu Glu Cys His Pro Leu Thr Met Thr Asp Pro Ile Glu Glu  
 245 250 255  
 His Arg Ile Cys Val Cys Val Arg Lys Arg Pro Leu Asn Lys Gln Glu  
 260 265 270  
 Leu Ala Lys Lys Glu Ile Asp Val Ile Ser Ile Pro Ser Lys Cys Leu  
 275 280 285  
 Leu Leu Val His Glu Pro Lys Leu Lys Val Asp Leu Thr Lys Tyr Leu  
 290 295 300  
 Glu Asn Gln Ala Phe Cys Phe Asp Phe Ala Phe Asp Glu Thr Ala Ser  
 305 310 315 320  
 Asn Glu Val Val Tyr Arg Phe Thr Ala Arg Pro Leu Val Gln Thr Ile  
 325 330 335  
 Phe Glu Gly Gly Lys Ala Thr Cys Phe Ala Tyr Gly Gln Thr Gly Ser  
 340 345 350  
 Gly Lys Thr His Thr Met Gly Gly Asp Leu Ser Gly Lys Ala Gln Asn  
 355 360 365  
 Ala Ser Lys Gly Ile Tyr Ala Met Ala Ser Arg Asp Val Phe Leu Leu  
 370 375 380  
 Lys Asn Gln Pro Cys Tyr Arg Lys Leu Gly Leu Glu Val Tyr Val Thr  
 385 390 395 400  
 Phe Phe Glu Ile Tyr Asn Gly Lys Leu Phe Asp Leu Leu Asn Lys Lys  
 405 410 415  
 Ala Lys Leu Arg Val Leu Glu Asp Gly Lys Gln Gln Val Gln Val Val  
 420 425 430  
 Gly Leu Gln Glu His Leu Val Asn Ser Ala Asp Asp Val Ile Lys Met  
 435 440 445  
 Leu Asp Met Gly Ser Ala Cys Arg Thr Ser Gly Gln Thr Phe Ala Asn  
 450 455 460  
 Ser Asn Ser Ser Arg Ser His Ala Cys Phe Gln Ile Ile Leu Arg Ala  
 465 470 475 480  
 Lys Gly Arg Met His Gly Lys Phe Ser Leu Val Asp Leu Ala Gly Asn  
 485 490 495  
 Glu Arg Gly Ala Asp Thr Ser Ser Ala Asp Arg Gln Thr Arg Met Glu

<400> 486						
gcgaaattga	ggtttcttgg	tattgcgcgt	ttctcttcct	tgctgactct	ccgaatggcc	60
atggactcgt	cgttcaggc	cgcctggtt	cccggtctcg	ctatcaagat	ccaacgcagt	120
aatggtttaa	ttcacagtgc	caatgtaagg	actgtgaact	tggagaaatc	ctgtgtttca	180
gtggaatggg	cagaaggagg	tgccacaaaag	ggcaaagaga	ttgattttga	tgatgtggct	240
gcaataaacc	cagaactctt	acagcttctt	cccttacatc	cgaaggacaa	tctgcccttg	300
caggaaaatc	taacaattcc	gaacaaaaa	cggagatccg	tcaactccaa	aattcctgct	360
ccaaaagaaa	gtcttcgaag	cgcgtccact	cgcattgtcca	ctgtctcaga	gcttcgcata	420
acggctcagg	agaatgacat	ggagggtggag	ctgcctgcag	ctgcaaacct	ccgaacgcag	480
ttttcagttc	ctctgcccc	cactaggcct	tctgccccg	cagtggctga	aataccattg	540
aggatggtca	gcgaggagat	ggaagagcaa	gtccattcca	tccgtggcag	ctcttctgca	600
aaccctgtga	actcagttcg	gaggaaatca	tgtcttgtga	aggaagtgga	aaaaatgaag	660
aacaagcgag	aagagagaag	ggcccagaa	tctgaaatga	gaatgaagag	agctcaggag	720
tatgacagta	gttttccaaa	ctgggaattt	gcccgaatga	ttaaagaatt	tccggctact	780
ttggaatgtc	atccactttac	tatgactgat	cctatcgaag	agcacagaat	atgtgtctgt	840
gttaggaaac	gccactgaa	taagcaagaa	ttggccaaga	agaaaattga	tgtgatttcc	900
attcctagca	agtgtctcct	cttggtacat	gaacccaagt	tgaagtggga	cttaacaaag	960
tatctggaga	accaagcatt	ctgctttgac	tttgcat ttg	atgaaacagc	ttcgaatgaa	1020
gttgtctaca	ggttcacagc	aaggccactg	gtacagacaa	tctttgaagg	tggaaaagca	1080
acttgttttg	catatggcca	gacaggaagt	ggcaagacac	atactatggg	cggagacctc	1140
tctgggaaag	cccagaatgc	atccaaaggg	atctatgcca	tggcctcccc	ggacgtcttc	1200
ctctcgaaag	atcaacctct	ctaccggaag	ttggccttg	aagtctatgt	gacattcttc	1260
gagatctaca	atgggaagct	gtttgacctg	ctcaacaaga	agggcaagct	gcgctgtctg	1320
gaggacggca	agcaacaggt	gcaagtgggtg	gggctgcagg	agcattctggt	taactgtgct	1380
gatgatgtca	tcaagatgct	cgacatgggc	agcgctgca	gaacctctgg	gcagacattt	1440



```

gccaaactcca attcctcccg ctcccacgcg tgcttccaaa ttattcttcg agctaaagg 1500
agaatgcatg gcaagtcttc tttggtagat ctggcagga atgagcgagg cgcagacact 1560
tccagtgtctg accggcagac ccgcatggag gggcagaaa tcaacaagag tctcttagcc 1620
ctgaaggagt gcatcagggc cctgggacag aacaaggctc acaccccggt cctgtgagagc 1680
aagctgacac aggtgtctgag ggactccttc attggggaga actctaggac ttgcatgatt 1740
gccacgatct caccaggcat aagctcctgt gaatatactt taaacaccct gagatatgca 1800
gacagggtca aggagctgag ccccccacagt gggcccagtg gagagcagtt gattcaaatg 1860
gaaacagaag agatggaagc ctgctctaac ggggcgctga ttccaggcaa tttatccaag 1920
gaagaggagg aactgtcttc ccagatgtcc agctttaacg aagccatgac tcagatcagg 1980
gagctggagg agaaggctat ggaagagctc aaggagatca tacagcaagg accagactgg 2040
cttgagctct ctgagatgac cgagcagcca gactatgacc tggagacctt tgtgaacaaa 2100
gcggaatctg ctctggccca gcaagccaag catttctcag ccctgcgaga tgtcatcaag 2160
gccttacgcc tggccatgca gctggaagag caggctagca gacaaataag cagcaagaaa 2220
cggccccagt gacgactgca aataaaaaatc tgtttggttt gacaccacagc ctcttccctg 2280
gccctcccca gagaactttg ggtacctggt gggcttaggc aggtctgag ctgggacagg 2340
ttctggtaaa tgccaagtat gggggcatct gggcccaggg cagctgggga gggggtcaga 2400
gtgacatggg acactccttt tctgttcttc agttgtcgcc ctcacgagag gaaggagctc 2460
ttagttaccc tttgtgttg ccttcttttc catcaagggg aatgttctca gcatagagct 2520
ttctccgag catctgcct gcgtggactg gctgctaata gagagctccc tgggggtgtc 2580
ctggctctgg ggagagagac ggagccttta gtacagctat ctgctggctc taaaccttct 2640
acgcctttgg gccgagcact gaatgtcttg tactttaaaa aaatgtttct gagacctctt 2700
tctactttac tgtctccta gactcctaga ggatccctac tgttttctgt tttatgtgtt 2760
tatacattgt atgtaacaat aaagagaaaa aataaaaaaa aaaaaaaaaa aaaaaaaaaa 2820
aaaaa 2825

```

&lt;210&gt; 487

&lt;211&gt; 566

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 487

```

Met Phe Val Ser Asp Phe Arg Lys Glu Phe Tyr Glu Val Val Gln Ser
  1           5           10           15
Gln Arg Val Leu Leu Phe Val Ala Ser Asp Val Asp Ala Leu Cys Ala
          20           25           30
Cys Lys Ile Leu Gln Ala Leu Phe Gln Cys Asp His Val Gln Tyr Thr
          35           40           45
Leu Val Pro Val Ser Gly Trp Gln Glu Leu Glu Thr Ala Phe Leu Glu
          50           55           60
His Lys Glu Gln Phe His Tyr Phe Ile Leu Ile Asn Cys Gly Ala Asn
        65           70           75           80
Val Asp Leu Leu Asp Ile Leu Gln Pro Asp Glu Asp Thr Ile Phe Phe
          85           90           95
Val Cys Asp Thr His Arg Pro Val Asn Val Val Asn Val Tyr Asn Asp
          100          105          110
Thr Gln Ile Lys Leu Leu Ile Lys Gln Asp Asp Asp Leu Glu Val Pro
          115          120          125
Ala Tyr Glu Asp Ile Phe Arg Asp Glu Glu Glu Asp Glu Glu His Ser
          130          135          140
Gly Asn Asp Ser Asp Gly Ser Glu Pro Ser Glu Lys Arg Thr Arg Leu
        145          150          155          160
Glu Glu Glu Ile Val Glu Gln Thr Met Arg Arg Arg Gln Arg Arg Glu
          165          170          175
Trp Glu Ala Arg Arg Arg Asp Ile Leu Phe Asp Tyr Glu Gln Tyr Glu
          180          185          190
Tyr His Gly Thr Ser Ser Ala Met Val Met Phe Glu Leu Ala Trp Met
          195          200          205
Leu Ser Lys Asp Leu Asn Asp Met Leu Trp Trp Ala Ile Val Gly Leu
          210          215          220
Thr Asp Gln Trp Val Gln Asp Lys Ile Thr Gln Met Lys Tyr Val Thr
        225          230          235          240

```

Asp Val Gly Val Leu Gln Arg His Val Ser Arg His Asn His Arg Asn  
 245 250 255  
 Glu Asp Glu Glu Asn Thr Leu Ser Val Asp Cys Thr Arg Ile Ser Phe  
 260 265 270  
 Glu Tyr Asp Leu Arg Leu Val Leu Tyr Gln His Trp Ser Leu His Asp  
 275 280 285  
 Ser Leu Cys Asn Thr Ser Tyr Thr Ala Ala Arg Phe Lys Leu Trp Ser  
 290 295 300  
 Val His Gly Gln Lys Arg Leu Gln Glu Phe Leu Ala Asp Met Gly Leu  
 305 310 315 320  
 Pro Leu Lys Gln Val Lys Gln Lys Phe Gln Ala Met Asp Ile Ser Leu  
 325 330 335  
 Lys Glu Asn Leu Arg Glu Met Ile Glu Glu Ser Ala Asn Lys Phe Gly  
 340 345 350  
 Met Lys Asp Met Arg Val Gln Thr Phe Ser Ile His Phe Gly Phe Lys  
 355 360 365  
 His Lys Phe Leu Ala Ser Asp Val Val Phe Ala Thr Met Ser Leu Met  
 370 375 380  
 Glu Ser Pro Glu Lys Asp Gly Ser Gly Thr Asp His Phe Ile Gln Ala  
 385 390 395 400  
 Leu Asp Ser Leu Ser Arg Ser Asn Leu Asp Lys Leu Tyr His Gly Leu  
 405 410 415  
 Glu Leu Ala Lys Lys Gln Leu Arg Ala Thr Gln Gln Thr Ile Ala Ser  
 420 425 430  
 Cys Leu Cys Thr Asn Leu Val Ile Ser Gln Gly Pro Phe Leu Tyr Cys  
 435 440 445  
 Ser Leu Met Glu Gly Thr Pro Asp Val Met Leu Phe Ser Arg Pro Ala  
 450 455 460  
 Ser Leu Ser Leu Leu Ser Lys His Leu Leu Lys Ser Phe Val Cys Ser  
 465 470 475 480  
 Thr Lys Asn Arg Arg Cys Lys Leu Leu Pro Leu Val Met Ala Ala Pro  
 485 490 495  
 Leu Ser Met Glu His Gly Thr Val Thr Val Val Gly Ile Pro Pro Glu  
 500 505 510  
 Thr Asp Ser Ser Asp Arg Lys Asn Phe Phe Gly Arg Ala Phe Glu Lys  
 515 520 525  
 Ala Ala Glu Ser Thr Ser Ser Arg Met Leu His Asn His Phe Asp Leu  
 530 535 540  
 Ser Val Ile Glu Leu Lys Ala Glu Asp Arg Ser Lys Phe Leu Asp Ala  
 545 550 555 560  
 Leu Ile Ser Leu Leu Ser  
 565

<210> 488  
 <211> 1938  
 <212> DNA  
 <213> Homo sapiens

<400> 488  
 gatttggcgg gagtcttgac cgccgccggg ctcttggtac ctcagcgcca gcgccaggcg 60  
 tccggccgcc gtggctatgt tcgtgtccga ttccgcgaaa gagttctacg aggtgggtcca 120  
 gagccagagg gtccttctct tcgtggcctc ggacgtggat gctctgtgtg cgtgcaagat 180  
 ccttcaggcc ttgttccagt gtgaccacgt gcaatatacg ctgggtccag tttctgggtg 240  
 gcaagaactt gaaactgcat ttcttgagca taaagaacag tticattatt ttattctcat 300  
 aaactgtgga gctaattgtag acctattgga tattcttcaa cctgatgaag acactatatt 360  
 ctttgtgtgt gacacccata ggccagtcaa tgcgtcaat gtatacaacg ataccagat 420  
 caaattactc attaaacaag atgatgacct tgaagttccc gcctatgaag acatcttcag 480  
 ggatgaagag gaggatgaag agcattcagg aaatgacagt gatgggtcag agccttctga 540  
 gaagcgca ca ggttagaag aggagatagt ggagcaaacc atgcggagga ggcagcggcg 600  
 agagtgggag gcccggagaa gagacatcct ctttgactac gagcagtatg aatatcatgg 660

```

gacatcggtca gccatggtga tgtttgagct ggcttggatg ctgtccaagg acctgaatga 720
catgctgtgg tgggcatcg ttggactaac agaccagtgg gtgcaagaca agatcactca 780
aatgaaatac gtgactgatg ttggtgtcct gcagcgccac gtttcccgcc acaaccaccg 840
gaacgaggat gaggagaaca cactctccgt ggactgcaca cggatctcct ttgagtatga 900
cctccgcctg gtgctctacc agcactgggc cctccatgac agcctgtgca acaccagcta 960
taccgcagcc aggttcaagc tgtggtctgt gcatggacag aagcggctcc aggagttcct 1020
tgcagacatg ggtcttcccc tgaagcaggt gaagcagaag ttccaggcca tggacatctc 1080
cttgaaggag aatttgcggt aaatgattga agagtctgca aataaatttg ggatgaagga 1140
catgcgcgtg cagactttca gcattcattt tgggttcaag cacaagtttc tggccagcga 1200
cgtggtcttt gccaccatgt ctttgatgga gagccccgag aaggatggct cagggacaga 1260
tcacttcac caggctctgg acagcctctc caggagtaac ctggacaagc tgtaccatgg 1320
cctggaactc gccaaagaagc agctgcgagc caccagcag accattgcca gctgcctttg 1380
caccaacctc gtcactctcc aggggccttt cctgtactgc tctctcatgg agggcactcc 1440
agatgtcatg ctgttctcta ggccggcatc cctaagcctg ctgagcaaac acctgctcaa 1500
gtcctttgtg tgttcgacaa agaaccggcg ctgcaaaactg ctgcccctgg tgatggctgc 1560
ccccctgagc atgggacatg gcacagtgc cgtggtgggc atccccccag agaccgacag 1620
ctcggacagg aagaactttt ttgggagggc gtttgagaag gcagcggaaa gcaccagctc 1680
ccgatgctg cacaaccatt ttgacctctc agtaattgag ctgaaagctg aggatcggag 1740
caagtttctg gacgcactta tttccctcct gtcctaggaa tttgattctt ccagaatgac 1800
cttcttattt atgtaactgg ctttcattta gattgtaagt tatggacatg atttgagatg 1860
tagaagccat tttttattaa ataaaatgct tatttttaggc tccgtcccca aaaaaaaaaa 1920
aaaaaaaaaa aaaaaaaaaa 1938

```

&lt;210&gt; 489

&lt;211&gt; 219

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 489

```

Met Ser Glu Thr Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Glu
1          5          10          15
Lys Thr Pro Val Lys Lys Lys Ala Arg Lys Ser Ala Gly Ala Ala Lys
20          25          30
Arg Lys Ala Ser Gly Pro Pro Val Ser Glu Leu Ile Thr Lys Ala Val
35          40          45
Ala Ala Ser Lys Glu Arg Ser Gly Val Ser Leu Ala Ala Leu Lys Lys
50          55          60
Ala Leu Ala Ala Ala Gly Tyr Asp Val Glu Lys Asn Asn Ser Arg Ile
65          70          75          80
Lys Leu Gly Leu Lys Ser Leu Val Ser Lys Gly Thr Leu Val Gln Thr
85          90          95
Lys Gly Thr Gly Ala Ser Gly Ser Phe Lys Leu Asn Lys Lys Ala Ala
100         105         110
Ser Gly Glu Ala Lys Pro Lys Ala Lys Lys Ala Gly Ala Ala Lys Ala
115         120         125
Lys Lys Pro Ala Gly Ala Ala Lys Lys Pro Lys Lys Ala Thr Gly Ala
130         135         140
Ala Thr Pro Lys Lys Ser Ala Lys Lys Thr Pro Lys Lys Ala Lys Lys
145         150         155         160
Pro Ala Ala Ala Ala Gly Ala Lys Lys Ala Lys Ser Pro Lys Lys Ala
165         170         175
Lys Ala Ala Lys Pro Lys Lys Ala Pro Lys Ser Pro Ala Lys Ala Lys
180         185         190
Ala Val Lys Pro Lys Ala Ala Lys Pro Lys Thr Ala Lys Pro Lys Ala
195         200         205
Ala Lys Pro Lys Lys Ala Ala Lys Lys Lys
210         215

```

&lt;210&gt; 490

&lt;211&gt; 785

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 490

```

agtccccggcc agtgcctctg cttccggctc gaattgctct cgctcacgct tgccttcaac 60
atgtccgaga ctgcgcctgc cgcgccgct gtcctggccc ctgccgagaa gactcccggtg 120
aagaagaagg cccgcaagtc tgcaggtgcg gccaaagcgca aagcgtcttg gcccccggtg 180
tccgagctca ttactaaagc tgttgccgcc tccaaggagc gcagcggcgt atctttggcc 240
gctctcaaga aagcgtggc agccgctggc tatgacgtgg agaagaacaa cagccgcctc 300
aagctgggtc tcaagagcct ggtgagcaag ggcaccctgg tgcagaccaa gggcaccggc 360
gcgtcgggtt ccttcaaact caacaagaag gcggcctctg gggaagccaa gcctaaggct 420
aaaaaggcag gcgcggccaa ggccaagaag ccagcaggag cggcgaagaa gcccagaag 480
gcgacggggg cggccacccc caagaagagc gccaaagaaga ccccaaagaa ggcgaagaag 540
ccggctgcag ctgctggagc caaaaaagcg aaaagcccga aaaaggcgaa agcagccaag 600
ccaaaaaagg cgcccaagag cccagcgaag gccaaagcag ttaaacccaa ggcgggctaaa 660
ccaaagaccg ccaagcccaa ggcagccaag ccaaagaagg cggcagccaa gaaaaagtag 720
aaagttcctt tggccaactg cttagaagcc caacacaacc caaaggctct tttcagagcc 780
accca

```

&lt;210&gt; 491

&lt;211&gt; 910

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 491

```

Met Arg Arg Ser Ala Ala Pro Ser Gln Leu Gln Gly Asn Ser Phe Lys
1      5      10      15
Lys Pro Lys Phe Ile Pro Pro Gly Arg Ser Asn Pro Gly Leu Asn Glu
20     25     30
Glu Ile Thr Lys Leu Asn Pro Asp Ile Lys Leu Phe Glu Gly Val Ala
35     40     45
Ile Asn Asn Thr Phe Leu Pro Ser Gln Asn Asp Leu Arg Ile Cys Ser
50     55     60
Leu Asn Leu Pro Ser Glu Glu Ser Thr Arg Glu Ile Asn Asn Arg Asp
65     70     75     80
Asn Cys Ser Gly Lys Tyr Cys Phe Glu Ala Pro Thr Leu Ala Thr Leu
85     90     95
Asp Pro Pro His Thr Val His Ser Ala Pro Lys Glu Val Ala Val Ser
100    105    110
Lys Glu Gln Glu Glu Lys Ser Asp Ser Leu Val Lys Tyr Phe Ser Val
115    120    125
Val Trp Cys Lys Pro Ser Lys Lys Lys His Lys Lys Trp Glu Gly Asp
130    135    140
Ala Val Leu Ile Val Lys Gly Lys Ser Phe Ile Leu Lys Asn Leu Glu
145    150    155    160
Gly Lys Asp Ile Gly Arg Gly Ile Gly Tyr Lys Phe Lys Glu Leu Glu
165    170    175
Lys Ile Glu Glu Gly Gln Thr Leu Met Ile Cys Gly Lys Glu Ile Glu
180    185    190
Val Met Gly Val Ile Ser Pro Asp Asp Phe Ser Ser Gly Arg Cys Phe
195    200    205
Gln Leu Gly Gly Gly Ser Thr Ala Ile Ser His Ser Ser Gln Val Ala
210    215    220
Arg Lys Cys Phe Ser Asn Pro Phe Lys Ser Val Cys Lys Pro Ser Ser
225    230    235    240
Lys Glu Asn Arg Gln Asn Asp Phe Gln Asn Cys Lys Pro Arg His Asp
245    250    255
Pro Tyr Thr Pro Asn Ser Leu Val Met Pro Arg Pro Asp Lys Asn His
260    265    270
Gln Trp Val Phe Asn Lys Asn Cys Phe Pro Leu Val Asp Val Val Ile
275    280    285

```

```

Asp Pro Tyr Leu Val Tyr His Leu Arg Pro His Gln Lys Glu Gly Ile
290                               295                 300
Ile Phe Leu Tyr Glu Cys Val Met Gly Met Arg Met Asn Gly Arg Cys
305                               310                 315                 320
Gly Ala Ile Leu Ala Asp Glu Met Gly Leu Gly Lys Thr Leu Gln Cys
                               325                 330                 335
Ile Ser Leu Ile Trp Thr Leu Gln Cys Gln Gly Pro Tyr Gly Gly Lys
                               340                 345                 350
Pro Val Ile Lys Lys Thr Leu Ile Val Thr Pro Gly Ser Leu Val Asn
                               355                 360                 365
Asn Trp Lys Lys Glu Phe Gln Lys Trp Leu Gly Ser Glu Arg Ile Lys
                               370                 375                 380
Ile Phe Thr Val Asp Gln Asp His Lys Val Glu Glu Phe Ile Lys Ser
385                               390                 395                 400
Ile Phe Tyr Ser Val Leu Ile Ile Ser Tyr Glu Met Leu Leu Arg Ser
                               405                 410                 415
Leu Asp Gln Ile Lys Asn Ile Lys Phe Asp Leu Leu Ile Cys Asp Glu
                               420                 425                 430
Gly His Arg Leu Lys Asn Ser Ala Ile Lys Thr Thr Thr Ala Leu Ile
                               435                 440                 445
Ser Leu Ser Cys Glu Lys Arg Ile Ile Leu Thr Gly Thr Pro Ile Gln
                               450                 455                 460
Asn Asp Leu Gln Glu Phe Phe Ala Leu Ile Asp Phe Val Asn Pro Gly
465                               470                 475                 480
Ile Leu Gly Ser Leu Ser Ser Tyr Arg Lys Ile Tyr Glu Glu Pro Ile
                               485                 490                 495
Ile Leu Ser Arg Glu Pro Ser Ala Ser Glu Glu Glu Lys Glu Leu Gly
                               500                 505                 510
Glu Arg Arg Ala Ala Glu Leu Thr Cys Leu Thr Gly Leu Phe Ile Leu
                               515                 520                 525
Arg Arg Thr Gln Glu Ile Ile Asn Lys Tyr Leu Pro Pro Lys Ile Glu
                               530                 535                 540
Asn Val Val Phe Cys Arg Pro Gly Ala Leu Gln Ile Glu Leu Tyr Arg
545                               550                 555                 560
Lys Leu Leu Asn Ser Gln Val Val Arg Phe Cys Leu Gln Gly Leu Leu
                               565                 570                 575
Glu Asn Ser Pro His Leu Ile Cys Ile Gly Ala Leu Lys Lys Leu Cys
                               580                 585                 590
Asn His Pro Cys Leu Leu Phe Asn Ser Ile Lys Glu Lys Glu Cys Ser
                               595                 600                 605
Ser Thr Cys Asp Lys Asn Glu Lys Ser Leu Tyr Lys Gly Leu Leu
                               610                 615                 620
Ser Val Phe Pro Ala Asp Tyr Asn Pro Leu Leu Phe Thr Glu Lys Glu
625                               630                 635                 640
Ser Gly Lys Leu Gln Val Leu Ser Lys Leu Leu Ala Val Ile His Glu
                               645                 650                 655
Leu Arg Pro Thr Glu Lys Val Val Leu Val Ser Asn Tyr Thr Gln Thr
                               660                 665                 670
Leu Asn Ile Leu Gln Glu Val Cys Lys Arg His Gly Tyr Ala Tyr Thr
                               675                 680                 685
Arg Leu Asp Gly Gln Thr Pro Ile Ser Gln Arg Gln Gln Ile Val Asp
690                               695                 700
Gly Phe Asn Ser Gln His Ser Ser Phe Phe Ile Phe Leu Leu Ser Ser
705                               710                 715                 720
Lys Ala Gly Gly Val Gly Leu Asn Leu Ile Gly Gly Ser His Leu Ile
                               725                 730                 735
Leu Tyr Asp Ile Asp Trp Asn Pro Ala Thr Asp Ile Gln Ala Met Ser
740                               745                 750
Arg Val Trp Arg Asp Gly Gln Lys Tyr Pro Val His Ile Tyr Arg Leu
755                               760                 765
Leu Thr Thr Gly Thr Ile Glu Glu Lys Ile Tyr Gln Arg Gln Ile Ser

```

770		775		780
Lys Gln Gly Leu Cys	Gly Ala Val Val Asp	Leu Thr Lys Thr Ser Glu		
785	790	795		800
His Ile Gln Phe Ser	Val Glu Glu Leu Lys	Asn Leu Phe Thr Leu His		
	805	810		815
Glu Ser Ser Asp Cys	Val Thr His Asp	Leu Leu Asp Cys Glu Cys Thr		
	820	825		830
Gly Glu Glu Val His	Thr Gly Asp Ser	Leu Glu Lys Phe Ile Val Ser		
	835	840		845
Arg Asp Cys Gln Leu	Gly Pro His His	Gln Lys Ser Asn Ser Leu Lys		
	850	855		860
Pro Leu Ser Met Ser	Gln Leu Lys Gln	Trp Lys His Phe Ser Gly Asp		
865	870	875		880
His Leu Asn Leu Thr	Asp Pro Phe Leu	Glu Arg Ile Thr Glu Asn Val		
	885	890		895
Ser Phe Ile Phe Gln	Asn Ile Thr Thr	Gln Ala Thr Gly Thr		
	900	905		910

<210> 492  
 <211> 3057  
 <212> DNA  
 <213> Homo sapiens

<400> 492  
 gagcttctcc ggcggcgagg ggatagctgg ttaccagaag gacttctttg cagggccagt 60  
 ggtttctgtc agatttttcgc cggtcacgac tgctgaatat gagacgatct gcagcaccaa 120  
 gtcagttgca ggggaattcc ttcaaaaaac caaaatttat acctccagga agaagtaatc 180  
 caggtctgaa tgaagagatt acaaaactga atccagatat aaaattattt gaggggtgtg 240  
 caattaataa cacctttctc ccgtcacaaa atgatcttag aatatgcagt ttaaatctgc 300  
 ctagtgaaga aagtactaga gaaatcaata acagagataa ttgcagtgga aaatattgtt 360  
 ttgaagcacc tacactggca acattagatc cacctcatat agttcattcg gctcctaaag 420  
 aagtagcagt gtccaaggaa caagaagaga aatctgatag cctagttaaa tatttcagtg 480  
 ttgtttggtg taagccttca aagaaaaaac ataaaaagtg ggaagggtgat gctgttctta 540  
 ttgtaaaagg aaagtcattt atattaaaga atttggaggg caaagacatt ggaagaggca 600  
 ttggttataa attcaaagag cttgaaaaga ttgaagaggg ccaaactctg atgatttgtg 660  
 gaaaagaaat agaagtcattg ggtgtaattc ctccagatga cttcagcagt ggcagggtgtt 720  
 ttacagcttg agagggaagt actgctatct cgcattcttc tcagggttgcc aggaaatgtt 780  
 tctctaacc tttcaaaagt gtttgtaaac caagttcaaa ttccctcggt atgccacgac 840  
 tccaaaattg caaaccacgc catgacccat atacgccaaa tccctcggt gatgtagtga 900  
 cagataagaa tcaccagtgg gtattcaata agaactgttt ccctcttggt gatgtagtga 960  
 ttgatcctta ccttgatatat catcttcgac cacatcagaa agaaggaatc atattccttt 1020  
 atgaatgtgt aatgggaatg agaatgaatg gcagatgtgg agctattctt gctgatgaaa 1080  
 tgggttttag gaagacattg caatgtattt cgctcatctg gaccctgcag tgcaggggac 1140  
 cctatggagg caagccagta ataaagaaga taggaagtga cactaattgt cacacctgga agcttgggtg 1200  
 ataattggaa gaaagaattt caaaaatggc taggaagtga aaggatcaag atatttactg 1260  
 ttgatcagga ccacaaagtt gaagaattca tcaagtctat attttattct gttcttatta 1320  
 tcagttatga aatgttactt cgttccctgg atcaaattaa gaatataaaa tttgatcttc 1380  
 taatctgtga cgaggggcat cgtttgaaaga acagtgccat taagacaact acagccctca 1440  
 ttagcctttc ttgtgagaaa agaataattc taactgggtac tccaattcag aatgatctgc 1500  
 aagaattttt tgcatttaatt gattttgtaa atccaggaat attaggtctt ttgtcatctt 1560  
 ataggaaaat atatgaagaa cccatcattt tatcgagaga accttctgct tctgaggaag 1620  
 aaaaggagtt aggagaaaga agagcagctg aacttacttg cctcactgga ctctttatcc 1680  
 ttagaagaac ccaagaaatt ataaataaat atctcccacc taaaatagag aatgttgctt 1740  
 ttgcccagacc aggagcacta cagattgagc ttatcgaaa gctgttaaat tctcaggttg 1800  
 tcaggttctg ccttcaaggg ttgttgaaaa atagtcacca tctaataatgt ataggagctc 1860  
 ttaaaaaact gtgcaatcac ccctgccttt tgttcaactc tataaaggaa aaggaaatgta 1920  
 gctcaacttg tgataaaaat gaagaaaaa gtcatacaaa aggcttgcta agtgtgtttc 1980  
 ctgctgacta caaccctctc ctgtttactg aaaggagtc aggaaaacta caggtgttgt 2040  
 ccaagctctt agcggttatc cacgaacttc gacctactga aaagggtgtg ttggtatcca 2100  
 actatacaca aaccttgaac atttacaag aagtatgtaa gcgtcatgga tatgcttata 2160

```

caagacttga tggacaaaca ccaatctctc aaaggcagca gattgttgat ggctttaaca 2220
gtcaacactc ttcttttttt atttttttgt taagtccaag agctgggtgt gtaggactta 2280
acctcattgg aggatctcac ttaattctct atgacattga ttggaatcca gccactgaca 2340
ttcaggcaat gtctagagta tggagagatg gtcagaaata tcctgtacat atttacagac 2400
tcctaactac aggtacaata gaagaaaaga tctatcaaag gcagatcagt aagcaagggtc 2460
tttgtggggc agttgtcgac ctccaaga catctgaaca tattcagttt tcagtagaag 2520
aacttaaaaa ttgtttcaca ttacatgaaa gttcagattg tgttactcat gatctgcttg 2580
actgtgagtg tacaggagaa gaagttcata caggtgattc gttggaaaaa ttcattgtct 2640
ctagagattg tcagcttggt ccacatcacc agaaatctaa ctccctgaaa cctctttcta 2700
gtgtccagct gaagcaatgg aaacattttt ctggagatca tttaaactct acagatcctt 2760
ttcttgaaag aataacagaa aatgtgtcat tcatTTTTtca gaataataacc actcaagcta 2820
ctggcacata gtgaaagatt acttctgaca ttccattgct cttcttttga aaattagtat 2880
ggtaattaaa tgtacttttt gaaaattaat agaattatTT aaattacagt atatgttgca 2940
aaatatatca cttttgatac aatagtcaaa attgagtggt ttaatgtttt gtaaatatta 3000
agtgtttaaa tgaaaaataa agatgtgctt atatcattgt aaaaaaaaaa aaaaaaa 3057

```

&lt;210&gt; 493

&lt;211&gt; 209

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 493

```

Met Gly Lys Gly Asp Pro Asn Lys Pro Arg Gly Lys Met Ser Ser Tyr
 1             5             10             15
Ala Phe Phe Val Gln Thr Cys Arg Glu Glu His Lys Lys Lys His Pro
      20             25             30
Asp Ser Ser Val Asn Phe Ala Glu Phe Ser Lys Lys Cys Ser Glu Arg
      35             40             45
Trp Lys Thr Met Ser Ala Lys Glu Lys Ser Lys Phe Glu Asp Met Ala
      50             55             60
Lys Ser Asp Lys Ala Arg Tyr Asp Arg Glu Met Lys Asn Tyr Val Pro
      65             70             75             80
Pro Lys Gly Asp Lys Lys Gly Lys Lys Lys Asp Pro Asn Ala Pro Lys
      85             90             95
Arg Pro Pro Ser Ala Phe Phe Leu Phe Cys Ser Glu His Arg Pro Lys
      100            105            110
Ile Lys Ser Glu His Pro Gly Leu Ser Ile Gly Asp Thr Ala Lys Lys
      115            120            125
Leu Gly Glu Met Trp Ser Glu Gln Ser Ala Lys Asp Lys Gln Pro Tyr
      130            135            140
Glu Gln Lys Ala Ala Lys Leu Lys Glu Lys Tyr Glu Lys Asp Ile Ala
      145            150            155            160
Ala Tyr Arg Ala Lys Gly Lys Ser Glu Ala Gly Lys Lys Gly Pro Gly
      165            170            175
Arg Pro Thr Gly Ser Lys Lys Lys Asn Glu Pro Glu Asp Glu Glu Glu
      180            185            190
Glu Glu Glu Glu Glu Asp Glu Asp Glu Glu Glu Asp Glu Asp Glu
      195            200            205
Glu

```

&lt;210&gt; 494

&lt;211&gt; 1277

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 494

```

gatgtggccc gtggcctagc tcgtcaagtt gccgtggcgc ggagaactct gcaaaacaag 60
aggctgagga ttgcgttaga gataaaccag ttcacgccgg agccccgtga gggaagcgct 120
tccgttgggt ccggccgctc tgcgggactc tgaggaaaag ctgcaccag gtggacgcgg 180

```

```

atctgtcaac atgggtaaag gagaccccaa caagccgagg ggcaaaatgt cctcgtacgc 240
cttcttcgtg cagacctgcc gggaagagca caagaagaaa caccgggact cttccgtcaa 300
tttcgcgga tttctccaaga agtggttcgga gagatggaag accatgtctg caaaggagaa 360
gtcgaagttt gaagatatgg caaaaagtga caaagctcgc tatgacaggg agatgaaaaa 420
ttacgttcct cccaaagggtg ataagaaggg gaagaaaaag gacccaatg ctcctaaaag 480
gccaccatct gccttcttcc tgttttctc tgaacatcgc ccaaagatca aaagtgaaca 540
ccctggccta tccattgggg atactgcaaa gaaattgggt gaaatgtggt ctgagcagtc 600
agccaaagat aaacaaccat atgaacagaa agcagctaag cttaaaggaga aatatgaaaa 660
ggatattgct gcatatcgtg ccaagggcaa aagtgaagca ggaaagaagg gccctggcag 720
gccaacaggc tcaaagaaga agaacgaacc agaagatgag gaggaggagg aggaagaaga 780
agatgaagat gaggaggaag aggatgaaga tgaagaataa atggctatcc tttaatgatg 840
cgtgtggaat gtgtgtgtgt gctcaggcaa ttattttgct aagaatgtga attcaagtgc 900
agctcaatac tagcttcagt ataaaaactg tacagatttt tgtatagctg ataagattct 960
ctgtagagaa aatactttta aaaaatgcag gttgtagctt tttgatgggc tactcataca 1020
gttagatttt acagcttctg atgttgaatg ttctaaata tttaatgggt tttttaattt 1080
ctgtgtgtat ggtagcacag caaacttgta ggaattagta tcaatagtaa attttgggtt 1140
ttttaggatg ttgcatttctg tttttttaa aaaaattttg taataaaatt atgtatatta 1200
tttctattgt ctttgtctta atatgctaag ttaattttca ctttaaaaaa gccatttgaa 1260
gacaaaaaaa aaaaaaa

```

1277

&lt;210&gt; 495

&lt;211&gt; 874

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 495

```

Met Asp Glu Glu Glu Asp Asn Leu Ser Leu Leu Thr Ala Leu Leu Glu
1      5      10      15
Glu Asn Glu Ser Ala Leu Asp Cys Asn Ser Glu Glu Asn Asn Phe Leu
20     25     30
Thr Arg Glu Asn Gly Glu Pro Asp Ala Phe Asp Glu Leu Phe Asp Ala
35     40     45
Asp Gly Asp Gly Glu Ser Tyr Thr Glu Glu Ala Asp Asp Gly Glu Thr
50     55     60
Gly Glu Thr Arg Asp Glu Lys Glu Asn Leu Ala Thr Leu Phe Gly Asp
65     70     75     80
Met Glu Asp Leu Thr Asp Glu Glu Glu Val Pro Ala Ser Gln Ser Thr
85     90     95
Glu Asn Arg Val Leu Pro Ala Pro Ala Pro Arg Arg Glu Lys Thr Asn
100    105    110
Glu Glu Leu Gln Glu Glu Leu Arg Asn Leu Gln Glu Gln Met Lys Ala
115    120    125
Leu Gln Glu Gln Leu Lys Val Thr Thr Ile Lys Gln Thr Ala Ser Pro
130    135    140
Ala Arg Leu Gln Lys Ser Pro Glu Lys Ser Pro Arg Pro Pro Leu Lys
145    150    155    160
Glu Arg Arg Val Gln Arg Ile Gln Glu Ser Thr Cys Phe Ser Ala Glu
165    170    175
Leu Asp Val Pro Ala Leu Pro Arg Thr Lys Arg Val Ala Arg Thr Pro
180    185    190
Lys Ala Ser Pro Pro Asp Pro Lys Ser Ser Ser Arg Met Thr Ser
195    200    205
Ala Pro Ser Gln Pro Leu Gln Thr Ile Ser Arg Asn Lys Pro Ser Gly
210    215    220
Ile Thr Arg Gly Gln Ile Val Gly Thr Pro Gly Ser Ser Gly Glu Thr
225    230    235    240
Thr Gln Pro Ile Cys Val Glu Ala Phe Ser Gly Leu Arg Leu Arg Arg
245    250    255
Pro Arg Val Ser Ser Thr Glu Met Asn Lys Lys Met Thr Gly Arg Lys
260    265    270
Leu Ile Arg Leu Ser Gln Ile Lys Glu Lys Met Ala Arg Glu Lys Leu

```



275					280					285					
Glu	Glu	Ile	Asp	Trp	Val	Thr	Phe	Gly	Val	Ile	Leu	Lys	Lys	Val	Thr
290						295					300				
Pro	Gln	Ser	Val	Asn	Ser	Gly	Lys	Thr	Phe	Ser	Ile	Trp	Lys	Leu	Asn
305				310						315					320
Asp	Leu	Arg	Asp	Leu	Thr	Gln	Cys	Val	Ser	Leu	Phe	Leu	Phe	Gly	Glu
			325						330					335	
Val	His	Lys	Ala	Leu	Trp	Lys	Thr	Glu	Gln	Gly	Thr	Val	Val	Gly	Ile
			340					345					350		
Leu	Asn	Ala	Asn	Pro	Met	Lys	Pro	Lys	Asp	Gly	Ser	Glu	Glu	Val	Cys
		355				360					365				
Leu	Ser	Ile	Asp	His	Pro	Gln	Lys	Val	Leu	Ile	Met	Gly	Glu	Ala	Leu
370						375				380					
Asp	Leu	Gly	Thr	Cys	Lys	Ala	Lys	Lys	Lys	Asn	Gly	Glu	Pro	Cys	Thr
385					390					395					400
Gln	Thr	Val	Asn	Leu	Arg	Asp	Cys	Glu	Tyr	Cys	Gln	Tyr	His	Val	Gln
			405					410					415		
Ala	Gln	Tyr	Lys	Lys	Leu	Ser	Ala	Lys	Arg	Ala	Asp	Leu	Gln	Ser	Thr
			420					425				430			
Phe	Ser	Gly	Gly	Arg	Ile	Pro	Lys	Lys	Phe	Ala	Arg	Arg	Gly	Thr	Ser
435						440					445				
Leu	Lys	Glu	Arg	Leu	Cys	Gln	Asp	Gly	Phe	Tyr	Tyr	Gly	Gly	Val	Ser
450						455				460					
Ser	Ala	Ser	Tyr	Ala	Ala	Ser	Ile	Ala	Ala	Ala	Val	Ala	Pro	Lys	Lys
465					470				475						480
Lys	Ile	Gln	Thr	Thr	Leu	Ser	Asn	Leu	Val	Val	Lys	Gly	Thr	Asn	Leu
			485					490					495		
Ile	Ile	Gln	Glu	Thr	Arg	Gln	Lys	Leu	Gly	Ile	Pro	Gln	Lys	Ser	Leu
		500						505					510		
Ser	Cys	Ser	Glu	Glu	Phe	Lys	Glu	Leu	Met	Asp	Leu	Pro	Thr	Cys	Gly
		515				520					525				
Ala	Arg	Asn	Leu	Lys	Gln	His	Leu	Ala	Lys	Ala	Thr	Ala	Ser	Gly	Ile
530						535				540					
Met	Gly	Ser	Pro	Lys	Pro	Ala	Ile	Lys	Ser	Ile	Ser	Ala	Ser	Ala	Leu
545				550					555						560
Leu	Lys	Gln	Gln	Lys	Gln	Arg	Met	Leu	Glu	Met	Arg	Arg	Arg	Lys	Ser
			565					570					575		
Glu	Glu	Ile	Gln	Lys	Arg	Phe	Leu	Gln	Ser	Ser	Ser	Glu	Val	Glu	Ser
			580					585				590			
Pro	Ala	Val	Pro	Ser	Ser	Ser	Arg	Gln	Pro	Pro	Ala	Gln	Pro	Pro	Arg
		595				600					605				
Thr	Gly	Ser	Glu	Phe	Pro	Arg	Leu	Glu	Gly	Ala	Pro	Ala	Thr	Met	Thr
610						615					620				
Pro	Lys	Leu	Gly	Arg	Gly	Val	Leu	Glu	Gly	Asp	Asp	Val	Leu	Phe	Tyr
625				630					635						640
Asp	Glu	Ser	Pro	Pro	Pro	Arg	Pro	Lys	Leu	Ser	Ala	Leu	Ala	Glu	Ala
			645					650					655		
Lys	Lys	Leu	Ala	Ala	Ile	Thr	Lys	Leu	Arg	Ala	Lys	Gly	Gln	Val	Leu
		660						665					670		
Thr	Lys	Thr	Asn	Pro	Asn	Ser	Ile	Lys	Lys	Lys	Gln	Lys	Asp	Pro	Gln
		675				680					685				
Asp	Ile	Leu	Glu	Val	Lys	Glu	Arg	Val	Glu	Lys	Asn	Thr	Met	Phe	Ser
690					695					700					
Ser	Gln	Ala	Glu	Asp	Glu	Leu	Glu	Pro	Ala	Arg	Lys	Lys	Arg	Arg	Glu
705				710					715						720
Gln	Leu	Ala	Tyr	Leu	Glu	Ser	Glu	Glu	Phe	Gln	Lys	Ile	Leu	Lys	Ala
			725					730					735		
Lys	Ser	Lys	His	Thr	Gly	Ile	Leu	Lys	Glu	Ala	Glu	Ala	Glu	Met	Gln
			740				745					750			
Glu	Arg	Tyr	Phe	Glu	Pro	Leu	Val	Lys	Lys	Glu	Gln	Met	Glu	Glu	Lys
755						760						765			

Met Arg Asn Ile Arg Glu Val Lys Cys Arg Val Val Thr Cys Lys Thr  
 770 775 780  
 Cys Ala Tyr Thr His Phe Lys Leu Leu Glu Thr Cys Val Ser Glu Gln  
 785 790 795 800  
 His Glu Tyr His Trp His Asp Gly Val Lys Arg Phe Phe Lys Cys Pro  
 805 810 815  
 Cys Gly Asn Arg Ser Ile Ser Leu Asp Arg Leu Pro Asn Lys His Cys  
 820 825 830  
 Ser Asn Cys Gly Leu Tyr Lys Trp Glu Arg Asp Gly Met Leu Lys Glu  
 835 840 845  
 Lys Thr Gly Pro Lys Ile Gly Gly Glu Thr Leu Leu Pro Arg Gly Glu  
 850 855 860  
 Glu His Ala Lys Phe Leu Asn Ser Leu Lys  
 865 870

<210> 496  
 <211> 4532  
 <212> DNA  
 <213> Homo sapiens

<400> 496

gaacgaagaa ggcgtcccg catcgccaa gattctacat tgctcatctg ggcattctgag 60  
 cctccttcga agtttctctg cacaactgtc ctcttgacag catggatgag gaggaagaca 120  
 atctgtctct gctgaccgca ctgctggaag aaaatgagtc agccttggat tgtaattcag 180  
 aagaaaataa cttcttgacg cgggaaaatg gcgagcccg cgcatttgat gagctctttg 240  
 atgccgacgg cgacgggtgaa tcttatacag aagaggctga tgatggagaa acaggagaga 300  
 caagagacga aaaggaaaat ctggccactc tctttggaga tatggaggac ttaacagatg 360  
 aagaagaagt tcccgcatca cagtcaactg aaaatagggg cctccctgct cctgccccca 420  
 ggcgagagaa aacgaatgaa gagttgcaag aggaattaag gaatttgcaa gagcaaatga 480  
 aggccttaca agagcagcta aaagtaacaa caattaaaca gacagcaagc ccagcccgtc 540  
 tgcaaaaatc ccctgagaag tctcccggc cacctcttaa ggagaggaga gttagagaga 600  
 ttcaggagtc aacatgcttt tctgcgagc ttgatgtccc tgcgtacca agaaccaaga 660  
 ggggtggctcg aacaccaaag gcttcacctc cagatcccaa aagctcatct tcaaggatga 720  
 caagtgcacc ctcccaaccc ctacagacga tttctcggaa caaacctagt gggataacta 780  
 gaggtcaaatt tgtggggacc ccaggaagtt ctggggaaac gactcaaccc atctgtgtgg 840  
 aagccttctc tggctctcgg ctcaggcggc ctcgagtatc ctccacagaa atgaacaaga 900  
 aaatgaccgg ccgaaaactg atcagactgt ctcagatcaa ggaaaagatg gccagagaga 960  
 agctggaaga aatagattgg gtgacatttg gggttatatt gaagaagggt acgccacaga 1020  
 gtgtgaatag tggaaaaacc ttcagcatat ggaaactgaa tgatctctgt gacctgacac 1080  
 aatgtgtgtc cttgttctta tttggagaag ttcacaaagc gctctggaag acggagcagg 1140  
 ggactgtcgt agggatcctc aatgccaacc ccatgaagcc caaggatggg tcagaggagg 1200  
 tgtgtttatc tatcgatcat cctcagaagg tcttaattat gggatgaagc cttgacctgg 1260  
 gaacctgtaa agccaagaag aagaatggag agcctgtcac gcagactgtg aatttgctgtg 1320  
 actgtgagta ctgtcagtac catgtccagg ctcatgacaa gaagctcagc gcaaagcgtg 1380  
 cggatctgca gtccaccttc tctggaggac gaattccaaa gagggttgcc cgcagaggca 1440  
 ccagcctcaa agaacggctg tgccaagatg gcttttacta cggagggggt tcttctgcct 1500  
 cgtatgcagc tccaattgca gcagctgtgg ctccaaagaa gaagattcaa accactctga 1560  
 gtaatctggt tgtaagggc acaaacttga tcatccagga aacacggcaa aaactcggaa 1620  
 taccacagaa gagcctgtct tgctctgagg agttcaagga actgatggac ctgccgacgt 1680  
 gtggagccag gaacttaaaa caacatttag ccaaagccac agcttcaggg attatgggga 1740  
 gcccaaaacc agccatcaag tccatctcgg cctcagcact cttgaagcaa cagaagcagc 1800  
 ggatgttgga gatgaggaga aggaaatcag aagaaatata gaagcgattt ctgcagagct 1860  
 caagtgaagt tgagagccca gctgtgccat cttcatcaag acagccccct gctcagcctc 1920  
 cacggacagg atccgagttc cccaggctgg agggagcccc ggccacaatg acgcccagc 1980  
 tggggcgagg tgtcttgaa ggagatgatg ttctctttta tgatgagtca ccaccacaa 2040  
 gaccaaaaact gagtgtctta gcagaagcca aaaagttagc tgctatcacc aaattaaggg 2100  
 caaaaggcca ggttcttaca aaaacaaacc caaacagcat taagaagaaa caaaaggacc 2160  
 ctcaggacat cctggagggtg aaggaaacgt tagaaaaaaa caccatgttt tcttctcaag 2220  
 ctgaggatga attggagcct gccaggaaaa aaaggagaga acaacttgcc tatctggaat 2280  
 ctgaggaatt tcagaaaatc ctaaaagcaa aatcaaaaca cacaggcatc ctgaaaggag 2340

```

ccgaggctga gatgcaggag cgctactttg agccactggg gaaaaaagaa caaatggaag 2400
aaaagatgag aaacatcaga gaagtgaagt gccgtgtcgt gacatgcaag acgtgcccct 2460
atacccaactt caagctgctg gagacctgcg tcagtgcgca gcatgaatac cactggcatg 2520
atgggtgtgaa gaggtttttc aaatgtccct gtggaaacag aagcatctcc ttggacagac 2580
tcccgaacaa gcactgcagt aactgtggcc tctacaaatg ggaacgggac ggaatgctaa 2640
agggaaaagac tggtccaaag ataggaggag aaactctgtt accaagagga gaagaacatg 2700
ctaaatttct gaacagcctt aaataacccg aacttcagac attttccac agacttcttg 2760
gcctcctgtg actctggaaa gcaaaggatt ggctgtgtat tgtccattga ttcttgattg 2820
acgccgtcaa aaacaaatgc ttgttaagcc cataagcttt gcctgcttac tttctgccat 2880
tgggttggtt tgataccaca tttaacattg acattttaagt ggaaaaccaa gttatcattg 2940
tcttttctaa gctcagtgtg gatgattgca ttacttcatt cactgaagtt tttgccccaa 3000
aattggaagg taaacagaga gctatgtttc tgtatctttt gggtatagag tgttcacttc 3060
tttatcataa caaaattcta gtgtttatac gaacacccag aggcaaaaga atttggttta 3120
attctcactc caggtaaagta gcttaacttc tgggcttcag ttttctcatc tgtaaaatca 3180
ggaagattgg actaagtgat cctgaaatgt attttttagc actggatttc taaaaataa 3240
aaaactttcc catctagata atgatgatca catagtcttg atgtacggac attaaaagcc 3300
agatttcttc attcaattct gttatctctg ttttactctt tgaaattgat caagccactg 3360
aatcactttg catttcagtt tatatatata gagagaaaga aggtgtctgc tcttacatta 3420
ttgtggagcc ctgtgataga aatatgtaaa atctcatatt attttttttt taattttttt 3480
attttttatg acaggggtctc actatgtcac cctggctgga gtgcagtagt gcgatcgcg 3540
cacactgcag ccttggtctc cctgggctca agcagtcctc ccacctcagt ctcccaata 3600
gtaggacta caggcgtgcg tgaccaagcc cagctaattt ttgcattttt ttagagatg 3660
gggttttgcc atgttgctca ggctggctc aaactcctga gcactagcaa tccaccacc 3720
tctgtttcca aaaaaaaaaa aaaaatgaaa ggtcaacccc tatgcaaatt accacagcaa 3780
aggtttcatt caggagattc tccatctgg gcaacctgg tttccaaata tcatttgacc 3840
taagtgaatg ttgatactag ctaaagattg ggtaaattgg ttgaattatt gtattgaagc 3900
ttgagctgta gctaaaagta atttaggttt cccctaagat gttattatgt tagggacata 3960
acacttttgg gaggttggtg tgggagatgg ttgatttagg ttttcaaaag ctagaaataa 4020
aatttacatg ccttagattt cataaaattc tgctctaatt gggtggaagg tgctgtatct 4080
aacttggtgt cctcctaagg ttatgtccta ataactattc ttttaggagt atacttctac 4140
tttatagaag gttgcttttc tttttaattt tttctaacia agaaaagaat aaagtattta 4200
ttaataagaa ccagaaagca cttgaaactg atgtttttaa tggctcattt agggtagatt 4260
tatttatctc attaacttaa aacagctatg tgtatgaaat aggtcacacac agaacttgaa 4320
caccaggttg gtgtctgagc aatccctttc ttatgggaaa aacaatgttc ttgtttgaac 4380
agagggatc attgcagtca gtattcacgt gtatattggt atataagttg tataatatgc 4440
ttgtaaaagg ctgaggtgag ctgtatctgg atgcctttt acaatttgat ttttaacttt 4500
aaaataaatt taaaacataa aaaaaaaaaa aa 4532

```

&lt;210&gt; 497

&lt;211&gt; 876

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 497

```

Met Glu Leu Ile Thr Ile Leu Glu Lys Thr Val Ser Pro Asp Arg Leu
 1           5           10           15
Glu Leu Glu Ala Ala Gln Lys Phe Leu Glu Arg Ala Ala Val Glu Asn
 20           25           30
Leu Pro Thr Phe Leu Val Glu Leu Ser Arg Val Leu Ala Asn Pro Gly
 35           40           45
Asn Ser Gln Val Ala Arg Val Ala Ala Gly Leu Gln Ile Lys Asn Ser
 50           55           60
Leu Thr Ser Lys Asp Pro Asp Ile Lys Ala Gln Tyr Gln Gln Arg Trp
 65           70           75           80
Leu Ala Ile Asp Ala Asn Ala Arg Arg Glu Val Lys Asn Tyr Val Leu
 85           90           95
Gln Thr Leu Gly Thr Glu Thr Tyr Arg Pro Ser Ser Ala Ser Gln Cys
100          105          110
Val Ala Gly Ile Ala Cys Ala Glu Ile Pro Val Asn Gln Trp Pro Glu
115          120          125
Leu Ile Pro Gln Leu Val Ala Asn Val Thr Asn Pro Asn Ser Thr Glu

```

130	135	140
His Met Lys Glu Ser Thr Leu Glu Ala Ile Gly Tyr Ile Cys Gln Asp		
145	150	155
Ile Asp Pro Glu Gln Leu Gln Asp Lys Ser Asn Glu Ile Leu Thr Ala		160
	165	170
Ile Ile Gln Gly Met Arg Lys Glu Glu Pro Ser Asn Asn Val Lys Leu		175
	180	185
Ala Ala Thr Asn Ala Leu Leu Asn Ser Leu Glu Phe Thr Lys Ala Asn		190
	195	200
Phe Asp Lys Glu Ser Glu Arg His Phe Ile Met Gln Val Val Cys Glu		205
	210	215
Ala Thr Gln Cys Pro Asp Thr Arg Val Arg Val Ala Ala Leu Gln Asn		220
225	230	235
Leu Val Lys Ile Met Ser Leu Tyr Tyr Gln Tyr Met Glu Thr Tyr Met		240
	245	250
Gly Pro Ala Leu Phe Ala Ile Thr Ile Glu Ala Met Lys Ser Asp Ile		255
	260	265
Asp Glu Val Ala Leu Gln Gly Ile Glu Phe Trp Ser Asn Val Cys Asp		270
	275	280
Glu Glu Met Asp Leu Ala Ile Glu Ala Ser Glu Ala Ala Glu Gln Gly		285
	290	295
Arg Pro Pro Glu His Thr Ser Lys Phe Tyr Ala Lys Gly Ala Leu Gln		300
305	310	315
Tyr Leu Val Pro Ile Leu Thr Gln Thr Leu Thr Lys Gln Asp Glu Asn		320
	325	330
Asp Asp Asp Asp Asp Trp Asn Pro Cys Lys Ala Ala Gly Val Cys Leu		335
	340	345
Met Leu Leu Ala Thr Cys Cys Glu Asp Asp Ile Val Pro His Val Leu		350
	355	360
Pro Phe Ile Lys Glu His Ile Lys Asn Pro Asp Trp Arg Tyr Arg Asp		365
	370	375
Ala Ala Val Met Ala Phe Gly Cys Ile Leu Glu Gly Pro Glu Pro Ser		380
385	390	395
Gln Leu Lys Pro Leu Val Ile Gln Ala Met Pro Thr Leu Ile Glu Leu		400
	405	410
Met Lys Asp Pro Ser Val Val Val Arg Asp Thr Ala Ala Trp Thr Val		415
	420	425
Gly Arg Ile Cys Glu Leu Leu Pro Glu Ala Ala Ile Asn Asp Val Tyr		430
	435	440
Leu Ala Pro Leu Leu Gln Cys Leu Ile Glu Gly Leu Ser Ala Glu Pro		445
	450	455
Arg Val Ala Ser Asn Val Cys Trp Ala Phe Ser Ser Leu Ala Glu Ala		460
465	470	475
Ala Tyr Glu Ala Ala Asp Val Ala Asp Asp Gln Glu Glu Pro Ala Thr		480
	485	490
Tyr Cys Leu Ser Ser Ser Phe Glu Leu Ile Val Gln Lys Leu Glu		495
	500	505
Thr Thr Asp Arg Pro Asp Gly His Gln Asn Asn Leu Arg Ser Ser Ala		510
	515	520
Tyr Glu Ser Leu Met Glu Ile Val Lys Asn Ser Ala Lys Asp Cys Tyr		525
	530	535
Pro Ala Val Gln Lys Thr Thr Leu Val Ile Met Glu Arg Leu Gln Gln		540
545	550	555
Val Leu Gln Met Glu Ser His Ile Gln Ser Thr Ser Asp Arg Ile Gln		560
	565	570
Phe Asn Asp Leu Gln Ser Leu Leu Cys Ala Thr Leu Gln Asn Val Leu		575
	580	585
Arg Lys Val Gln His Gln Asp Ala Leu Gln Ile Ser Asp Val Val Met		590
	595	600
Ala Ser Leu Leu Arg Met Phe Gln Ser Thr Ala Gly Ser Gly Gly Val		605
610	615	620

Gln Glu Asp Ala Leu Met Ala Val Ser Thr Leu Val Glu Val Leu Gly  
 625 630 635 640  
 Gly Glu Phe Leu Lys Tyr Met Glu Ala Phe Lys Pro Phe Leu Gly Ile  
 645 650 655  
 Gly Leu Lys Asn Tyr Ala Glu Tyr Gln Val Cys Leu Ala Ala Val Gly  
 660 665 670  
 Leu Val Gly Asp Leu Cys Arg Ala Leu Gln Ser Asn Ile Ile Pro Phe  
 675 680 685  
 Cys Asp Glu Val Met Gln Leu Leu Glu Asn Leu Gly Asn Glu Asn  
 690 695 700  
 Val His Arg Ser Val Lys Pro Gln Ile Leu Ser Val Phe Gly Asp Ile  
 705 710 715 720  
 Ala Leu Ala Ile Gly Gly Glu Phe Lys Lys Tyr Leu Glu Val Val Leu  
 725 730 735  
 Asn Thr Leu Gln Ala Ser Gln Ala Gln Val Asp Lys Ser Asp Tyr  
 740 745 750  
 Asp Met Val Asp Tyr Leu Asn Glu Leu Arg Glu Ser Cys Leu Glu Ala  
 755 760 765  
 Tyr Thr Gly Ile Val Gln Gly Leu Lys Gly Asp Gln Glu Asn Val His  
 770 775 780  
 Pro Asp Val Met Leu Val Gln Pro Arg Val Glu Phe Ile Leu Ser Phe  
 785 790 795 800  
 Ile Asp His Ile Ala Gly Asp Glu Asp His Thr Asp Gly Val Val Ala  
 805 810 815  
 Cys Ala Ala Gly Leu Ile Gly Asp Leu Cys Thr Ala Phe Gly Lys Asp  
 820 825 830  
 Val Leu Lys Leu Val Glu Ala Arg Pro Met Ile His Glu Leu Leu Thr  
 835 840 845  
 Glu Gly Arg Arg Ser Lys Thr Asn Lys Ala Lys Thr Leu Ala Thr Trp  
 850 855 860  
 Ala Thr Lys Glu Leu Arg Lys Leu Lys Asn Gln Ala  
 865 870 875

&lt;210&gt; 498

&lt;211&gt; 4205

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 498

```

ctccctcgct ccctccctgc gcgcgcctc tcactcacag cctcccttcc ttcttttctcc 60
ctccgcctcc cgagcaccag cgcgctctga gctgccccca gggctccctcc cccgcgcgcca 120
gcagccatt tggaggagg aagtaaggga agaggagagg aaggggagcc ggaccgacta 180
ccagacaga gccggtgaat gggtttgtgg tgaccccgcc cccccacccc accctccctt 240
cccacccgac ccccaacccc catccccagt tcgagccgcc gcccgaaagg ccggggccgtc 300
gtcttaggag gagtgcgcgc cgccgccacc tccgccatgg agctgatcac cattctcgag 360
aagaccgtgt ctcccgatcg gctggagctg gaagcggcgc agaagttcct ggagcgtgcg 420
gccgtggaga acctgcccac tttccttgtg gaactgtcca gagtgtctggc aaatccagga 480
aacagtcagg ttgccagagt tgcagctggg ctacaaatca agaactcttt gacatctaaa 540
gatccagata tcaaggcaca atatcagcag aggtggcttg ctattgatgc taatgctcga 600
cgagaagtca agaactatgt tttgcagaca ttgggtacag aaacttaccg gcctagtctt 660
gcctcacagt gtgtggctgg tattgcttgt gcagagatcc cagtaaacca gtggccagaa 720
ctcattctc agctgggtgg caatgtcaca aacccaaca gcacagagca catgaaggag 780
tcgacattgg aagccatcgg ttatatttgc caagatatag acccagagca gctacaagat 840
aatccaatg agattctgac tgccataatc caggggatga ggaaagaaga gcctagtaat 900
aatgtgaagc tagctgctac gaatgcactc ctgaactcat tggagttcac caaagcaaac 960
tttgataaag agtctgaaag gcactttatt atgcaggtgg tctgtgaagc cacacagtg 1020
ccagatacga gggtagaggt ggctgcttta cagaatctgg tgaagataat gtccttatat 1080
tatcagtaca tggagacata tatgggtcct gctctttttg caatcacaat cgaagcaatg 1140
aaaagtgaca ttgatgaggt ggctttacaa gggatagaat tctgggtcaa tgtctgtgat 1200
gaggaaatgg atttggccat tgaagcttca gaggcagcag aacaaggacg gccccctgag 1260

```

cacaccagca agttttatgc gaaggagca ctacagtatc tggttccaat cctcacacag 1320  
 acactaacta aacaggacga aaatgatgat gacgatgact ggaacccctg caaagcagca 1380  
 ggggtgtgcc tcatgcttct ggccacctgc tgtgaagatg acattgtccc acatgtcctc 1440  
 cccttcatta aagaacacat caagaaccca gattggcggt accgggatgc agcagtgatg 1500  
 gcttttgggt gtatcttggg aggaccagag ccgagtcagc tcaaaccact agttatacag 1560  
 gctatgcccc ccctaataga attaatgaaa gaccccgatg tagttgttcg agatacagct 1620  
 gcatggactg taggcagaat ttgtgagctg ctctctgaag ctgccatcaa tgatgtctac 1680  
 ttggctcccc tgctacagtg tctgattgag ggtctcagtg ctgaaccag agtggttca 1740  
 aatgtgtgct gggctttctc cagtctggct gaagctgctt atgaagctgc agacgttgct 1800  
 gatgatcagg aagaaccagc tacttactgc ttatcttctt catttgaact catagtccag 1860  
 aagctcctag agactacaga cagacctgat ggacaccaga acaacctgag gagttctgca 1920  
 tatgaatctc tgatggaaat tgtgaaaaac cgtgccaagg attgttatcc tgctgtccag 1980  
 aaaacgactt tggctcatcat ggaacgactg caacaggttc ttcagatgga gtcacatctc 2040  
 cagagcacat ccgatagaat ccagttcaat gaccttcagt ctttactctg tgcaactctt 2100  
 cagaatgttc ttccgaaagt gcaacatcaa gatgctttgc agatctctga tgtggttatg 2160  
 gcctccctgt taaggatggt ccaaagcaca gctgggtctg ggggagtaca agaggatgcc 2220  
 ctgatggcag ttagcacact ggtggaagtg ttgggtggtg aattccctcaa gtacatggag 2280  
 gcctttaaac ccttctggg cattggatta aaaaattatg ctgaatacca ggtttgtttg 2340  
 gcagctgtgg gcttagtggt agacttggtc cgtgccctgc aatccaacat cataccttct 2400  
 tgtgacgagg tgatgcagct gcttctggaa aatttgggga atgagaacgt ccacaggtct 2460  
 gtgaagccgc agattctgtc agtgtttggt gatattgccc ttgctattgg aggagagttt 2520  
 aaaaaatact tagagttgt attgaatact cttcagcagg cctccaagc ccaggtggag 2580  
 aagtcagact atgacatggt ggattatctg aatgagctaa gggaaagctg cttggaagcc 2640  
 tatactggaa tcgtccaggg attaaagggg gatcaggaga acgtacaccc ggatgtgatg 2700  
 ctggtacaac ccagagtga atttattctg tctttcattg accacattgc tggagatgag 2760  
 gatcacacag atggagtagt agcttgtgct gctggactaa taggggactt atgtacagca 2820  
 tttgggaagg atgtactgaa attagtagaa gctaggccaa tgatccatga attgttaact 2880  
 gaagggcgga gatcgaagac taacaaagca aaaacccttg ctacatgggc aacaaaagaa 2940  
 ctgaggaaac tgaagaacca agcttgatct gttaccattg ggatgataac ctgaggaccc 3000  
 ccactggaaa tctcccatct tttgaaaac ctggaagtga ggagtgtgca cggatgctga 3060  
 atgtttggga atgagaggat gagtgagtga ggcttgaaa cacaccacat tgaaaatcct 3120  
 gccacagcag cagccgcagc cgccaacagc agcgtgtta gtgagctaag taagcactga 3180  
 cttcgtagaa aaccataaca tcggccatct tggaaaagag aaaaacaatg gagttaacta 3240  
 tttaaaaaaa aagaaagaaa gttatctctt cccaggagag gctagaagta gcttttctgt 3300  
 cttttggcca gtgccgagt gaatgcctgg tttgggggag gaggagggac tgggttcagc 3360  
 tgtggtgctt tgttgtaaaa ggcagcctgg cctttgctac tgaggagaaa gatggagcct 3420  
 ggggtctcaag cccacctctg ctgtaccttt gccacatggt actgtatgct tgccagctag 3480  
 aaggagggtg agggattttt tacagtctga gaatgagtgt gtgtgagtga ggcggtatcc 3540  
 acatttctca cttcaagtca ttgcagtttc tttttcccag aaaacaagg gttagatgtt 3600  
 gcatttcata aaactaaccg aagttctgtc tactgatgca gcacaagaga tgtaaaaaaa 3660  
 aaaaaaaaaa aaaaaaaaaa aacacacaca cagaggaaag acgctcttta ggttttgttt 3720  
 tgtttttttt ttttggtttt gttttttgtt ttttttactc tagggaaaac actgacgaat 3780  
 ggtcagagct cctatcctga tcttttcatc aaggcgctt tcctaataat atggttcaac 3840  
 tgtgaatgta gaagtgggg ggagggggga gaaaaagaaa actctggcgt tagaggatat 3900  
 agaaaaatat aagtacaatt gttacaaata acgcagactt caaaaacaaa aaaatcacia 3960  
 cccaaacaaa ccaaaattta aatgatcaga attggcagca caaagaaaac gccctctcct 4020  
 gacttgattt gtggcagctt gaacgcccc agaaaattgt gccaaagagt ttagaaaaat 4080  
 aaatatacaa taaaagtaaa cacatacaca caaacagca aacttcaggt aactattttg 4140  
 gattgcaaac aggataaatt aaatgttcaa acaatctgat aaaataacca tttggaaact 4200  
 gaaaa 4205

&lt;210&gt; 499

&lt;211&gt; 763

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 499

Met Lys Ala Ser Pro Arg Arg Pro Leu Ile Leu Lys Arg Arg Arg Leu  
 1 5 10 15  
 Pro Leu Pro Val Gln Asn Ala Pro Ser Glu Thr Ser Glu Glu Glu Pro  
 20 25 30

Lys Arg Ser Pro Ala Gln Gln Glu Ser Asn Gln Ala Glu Ala Ser Lys  
 35 40 45  
 Glu Val Ala Glu Ser Asn Ser Cys Lys Phe Pro Ala Gly Ile Lys Ile  
 50 55 60  
 Ile Asn His Pro Thr Met Pro Asn Thr Gln Val Val Ala Ile Pro Asn  
 65 70 75 80  
 Asn Ala Asn Ile His Ser Ile Ile Thr Ala Leu Thr Ala Lys Gly Lys  
 85 90 95  
 Glu Ser Gly Ser Ser Gly Pro Asn Lys Phe Ile Leu Ile Ser Cys Gly  
 100 105 110  
 Gly Ala Pro Thr Gln Pro Pro Gly Leu Arg Pro Gln Thr Gln Thr Ser  
 115 120 125  
 Tyr Asp Ala Lys Arg Thr Glu Val Thr Leu Glu Thr Leu Gly Pro Lys  
 130 135 140  
 Pro Ala Ala Arg Asp Val Asn Leu Pro Arg Pro Pro Gly Ala Leu Cys  
 145 150 155 160  
 Glu Gln Lys Arg Glu Thr Cys Ala Asp Gly Glu Ala Ala Gly Cys Thr  
 165 170 175  
 Ile Asn Asn Ser Leu Ser Asn Ile Gln Trp Leu Arg Lys Met Ser Ser  
 180 185 190  
 Asp Gly Leu Gly Ser Arg Ser Ile Lys Gln Glu Met Glu Glu Lys Glu  
 195 200 205  
 Asn Cys His Leu Glu Gln Arg Gln Val Lys Val Glu Glu Pro Ser Arg  
 210 215 220  
 Pro Ser Ala Ser Trp Gln Asn Ser Val Ser Glu Arg Pro Pro Tyr Ser  
 225 230 235 240  
 Tyr Met Ala Met Ile Gln Phe Ala Ile Asn Ser Thr Glu Arg Lys Arg  
 245 250 255  
 Met Thr Leu Lys Asp Ile Tyr Thr Trp Ile Glu Asp His Phe Pro Tyr  
 260 265 270  
 Phe Lys His Ile Ala Lys Pro Gly Trp Lys Asn Ser Ile Arg His Asn  
 275 280 285  
 Leu Ser Leu His Asp Met Phe Val Arg Glu Thr Ser Ala Asn Gly Lys  
 290 295 300  
 Val Ser Phe Trp Thr Ile His Pro Ser Ala Asn Arg Tyr Leu Thr Leu  
 305 310 315 320  
 Asp Gln Val Phe Lys Pro Leu Asp Pro Gly Ser Pro Gln Leu Pro Glu  
 325 330 335  
 His Leu Glu Ser Gln Gln Lys Arg Pro Asn Pro Glu Leu Arg Arg Asn  
 340 345 350  
 Met Thr Ile Lys Thr Glu Leu Pro Leu Gly Ala Arg Arg Lys Met Lys  
 355 360 365  
 Pro Leu Leu Pro Arg Val Ser Ser Tyr Leu Val Pro Ile Gln Phe Pro  
 370 375 380  
 Val Asn Gln Ser Leu Val Leu Gln Pro Ser Val Lys Val Pro Leu Pro  
 385 390 395 400  
 Leu Ala Ala Ser Leu Met Ser Ser Glu Leu Ala Arg His Ser Lys Arg  
 405 410 415  
 Val Arg Ile Ala Pro Lys Val Leu Leu Ala Glu Glu Gly Ile Ala Pro  
 420 425 430  
 Leu Ser Ser Ala Gly Pro Gly Lys Glu Glu Lys Leu Leu Phe Gly Glu  
 435 440 445  
 Gly Phe Ser Pro Leu Leu Pro Val Gln Thr Ile Lys Glu Glu Glu Ile  
 450 455 460  
 Gln Pro Gly Glu Glu Met Pro His Leu Ala Arg Pro Ile Lys Val Glu  
 465 470 475 480  
 Ser Pro Pro Leu Glu Glu Trp Pro Ser Pro Ala Pro Ser Phe Lys Glu  
 485 490 495  
 Glu Ser Ser His Ser Trp Glu Asp Ser Ser Gln Ser Pro Thr Pro Arg  
 500 505 510  
 Pro Lys Lys Ser Tyr Ser Gly Leu Arg Ser Pro Thr Arg Cys Val Ser

515	520	525
Glu Met Leu Val Ile Gln His Arg Glu Arg Arg Glu Arg Ser Arg Ser		
530	535	540
Arg Arg Lys Gln His Leu Pro Pro Cys Val Asp Glu Pro Glu Leu		
545	550	555
Leu Phe Ser Glu Gly Pro Ser Thr Ser Arg Trp Ala Ala Glu Leu Pro		
565	570	575
Phe Pro Ala Asp Ser Ser Asp Pro Ala Ser Gln Leu Ser Tyr Ser Gln		
580	585	590
Glu Val Gly Gly Pro Phe Lys Thr Pro Ile Lys Glu Thr Leu Pro Ile		
595	600	605
Ser Ser Thr Pro Ser Lys Ser Val Leu Pro Arg Thr Pro Glu Ser Trp		
610	615	620
Arg Leu Thr Pro Pro Ala Lys Val Gly Gly Leu Asp Phe Ser Pro Val		
625	630	635
Gln Thr Ser Gln Gly Ala Ser Asp Pro Leu Pro Asp Pro Leu Gly Leu		
645	650	655
Met Asp Leu Ser Thr Thr Pro Leu Gln Ser Ala Pro Pro Leu Glu Ser		
660	665	670
Pro Gln Arg Leu Leu Ser Ser Glu Pro Leu Asp Leu Ile Ser Val Pro		
675	680	685
Phe Gly Asn Ser Ser Pro Ser Asp Ile Asp Val Pro Lys Pro Gly Ser		
690	695	700
Pro Glu Pro Gln Val Ser Gly Leu Ala Ala Asn Arg Ser Leu Thr Glu		
705	710	715
Gly Leu Val Leu Asp Thr Met Asn Asp Ser Leu Ser Lys Ile Leu Leu		
725	730	735
Asp Ile Ser Phe Pro Gly Leu Asp Glu Asp Pro Leu Gly Pro Asp Asn		
740	745	750
Ile Asn Trp Ser Gln Phe Ile Pro Glu Leu Gln		
755	760	

&lt;210&gt; 500

&lt;211&gt; 3336

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (3317)...(3317)

&lt;223&gt; N=A, T, G, or C

&lt;400&gt; 500

```

cggcggcgac tgcagtctgg aggggtccaca cttgtgattc tcaatggaga gtgaaaacgc 60
agattcataa tgaaagctag ccccgctcgg ccactgattc tcaaaagacg gaggtgccc 120
cttcctgttc aaaatgcccc aagtgaacaa tcagaggagg aacctaaagag atccccctgcc 180
caacaggagt ctaatcaagc agaggcctcc aagggaagtgg cggagtccaa ctcttgcaag 240
tttcagctg ggatcaagat tattaaccac cccaccatgc ccaacacgca agtagtggcc 300
atccccaaca atgctaatat tcacagcadc atcacagcac tgactgccaa gggaaaagag 360
agtggcagta gtgggcccac caaattcatc ctcacagct gtggggggagc cccaactcag 420
cctccaggac tccggcctca aacccaaacc agctatgatg ccaaaaggac agaagtgacc 480
ctggagacct tgggaccaa acctgcagct agggatgtga atcttcctag accacctgga 540
gccctttgcg agcagaaacg ggagacctgt gcagatggtg aggcagcagg ctgcactatc 600
aacaatagcc tatccaacat ccagtggcct cgaaagatga gttctgatgg actgggctcc 660
cgcagcatca agcaagagat ggaggaaaag gagaattgtc acctggagca gcgacaggtt 720
aagggttgagg agccttcgag accatcagcg tcctggcaga actctgtgtc tgagcggcca 780
ccctactctt acatggccat gatacaattc gccatcaaca gcactgagag gaagcgcagt 840
actttgaaag acatctatac gtggattgag gaccactttc cctactttta gcacattgcc 900
aagccaggct ggaagaactc catccgccac aacctttccc tgcacgacat gttgtccgg 960
gagacgtctg ccaatggcaa ggtctccttc tggaccattc accccagtgc caaccgctac 1020

```



```

ttgacattgg accaggtggt taagccactg gaccagggtg ctccacaatt gcccagacac 1080
ttggaatcac agcagaaacg accgaatcca gagctccgcc ggaacatgac catcaaaacc 1140
gaactcccc tgggcgcacg gcggaagatg aagccactgc taccacgggt cagctcatal 1200
ctggtacctt tccagttccc ggtgaaccag tcaactggtg tgcagccctc ggtgaagggt 1260
ccattgcccc tgggcgcttc cctcatgagc tcagagcttg cccgccatag caagcgagtc 1320
cgcattgccc ccaaggtgct gctagctgag gaggggatag ctctcttttc ttctgcagga 1380
ccagggaag aggaagaaact cctgtttgga gaagggtttt ctcttttgct tccagttcag 1440
actatcaagg aggaagaaat ccagcctggg gaggaatgc cacacttagc gagacccatc 1500
aaagtggaga gccctccctt ggaagagtgg cctcccccg cccatctttt caaagaggaa 1560
tcatctcact cctgggagga ttctgcccaa tctcccccc caagacccaa gaagtcctac 1620
agtgggctta ggtccccaac ccggtgtgtc tcggaatgc ttgtgattca acacagggtg 1680
aggagggaga ggagccggtc tcggaggaaa cagcatctac tgcctccctg tgtggatgag 1740
ccggagctgc tcttctcaga ggggccaggt acttcccgct gggccgcaga gctcccggtc 1800
ccagcagact cctctgaccc tgctctccag ctcaactact cccaggaagt gggaggacct 1860
tttaagacac ccattaagga aacgctgccc atctctccca ccccgagcaa atctgtcttc 1920
cccgaacccc ctgaatcctg gaggtcaccg cccccagcca aagtaggggg actggatttc 1980
agcccagtac aaacctccca ggggtgctct cagcccttgc ctgacccctt ggggtgatg 2040
gatctcagca ccaactccctt gcaaagtgtc ccccccttg aatcacgcga aaggctcttc 2100
agttcagaac ccttagacct catctccgtc ccttttgga actcttctcc ctcatatata 2160
gacgtcccca agccaggtc cccggagcca caggtttctg gccttgagc caatcgttct 2220
ctgacagaag gcctggtcct ggacacaatg aatgacagcc tcagcaagat cctgctggac 2280
atcagcttct ctggcctgga cgaggaccca ctgggcccctg acaacatcaa ctggtcccag 2340
tttattcctg agctacagta gagccctgcc ctggcccctg tgctcaagct gtccaccatc 2400
ccgggcactc caaggctcag tgcaccccaa gcctctgagt gaggacagca ggcagggtg 2460
gttctgtctc tcatagctcc ctgctgctg attatgcaa agtagcagtc acaccctagc 2520
cactgctggg acctgtgtt ccccaagagt atctgattcc tctgtgtcc ctgccaggag 2580
ctgaagggtg ggaacaacaa aggcaatggt gaaaagagat taggaacccc ccagcctgtt 2640
tccattctct gccagcaggt ctcttacctt cctgatctt tgcagggtgg tccgtgtaaa 2700
tagtataaat tctccaaatt atctctaat tataaatgta agcttatttc cttagatcat 2760
tatccagaga ctgccagaag gtgggtagga tgacctggg tttcaattga cttctgttcc 2820
ttgcttttag tttgataga agggaagacc tgcagtgcac ggtttcttcc aggctgaggt 2880
acctggatct tgggttcttc actgcaggga cccagacaag tggatctgct tgccagagtc 2940
ctttttgccc ctccctgcca cctccccgtg tttccaagtc agctttctg caagaagaaa 3000
tcttggttaa aaaagtcttt tgtattgggt caggagttga atttgggggtg ggaggatgga 3060
tgcaactgaa gcagagtgtg ggtgcccaga tgtgcgctat tagatgttct tctgataatg 3120
tccccaatca taccaggag actggcattg acgagaactc aggtggaggc ttgagaaggc 3180
cgaaagggcc cctgacctgc ctggcttctt ctacttgccc ctgagcttg caaagagcca 3240
ccctaggccc cagctgaccg catgggtgtg agccagcttg agaacactaa ctactcaata 3300
aaagcgaagg tggaccnaaa aaaaaaaaaa aaaaaa 3336

```

&lt;210&gt; 501

&lt;211&gt; 952

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 501

```

Met Ala Glu Gly Gly Ala Ala Asp Leu Asp Thr Gln Arg Ser Asp Ile
  1           5           10          15
Ala Thr Leu Leu Lys Thr Ser Leu Arg Lys Gly Asp Thr Trp Tyr Leu
          20          25          30
Val Asp Ser Arg Trp Phe Lys Gln Trp Lys Lys Tyr Val Gly Phe Asp
          35          40          45
Ser Trp Asp Lys Tyr Gln Met Gly Asp Gln Asn Val Tyr Pro Gly Pro
          50          55          60
Ile Asp Asn Ser Gly Leu Leu Lys Asp Gly Asp Ala Gln Ser Leu Lys
          65          70          75          80
Glu His Leu Ile Asp Glu Leu Asp Tyr Ile Leu Leu Pro Thr Glu Gly
          85          90          95
Trp Asn Lys Leu Val Ser Trp Tyr Thr Leu Met Glu Gly Gln Glu Pro
          100         105         110
Ile Ala Arg Lys Val Val Glu Gln Gly Met Phe Val Lys His Cys Lys

```

115	120	125
Val. Glu Val Tyr Leu Thr Glu Leu Lys Leu Cys Glu Asn Gly Asn Met		
130	135	140
Asn Asn Val Val Thr Arg Phe Ser Lys Ala Asp Thr Ile Asp Thr		
145	150	155
Ile Glu Lys Glu Ile Arg Lys Ile Phe Ser Ile Pro Asp Glu Lys Glu		
165	170	175
Thr Arg Leu Trp Asn Lys Tyr Met Ser Asn Thr Phe Glu Pro Leu Asn		
180	185	190
Lys Pro Asp Ser Thr Ile Gln Asp Ala Gly Leu Tyr Gln Gly Gln Val		
195	200	205
Leu Val Ile Glu Gln Lys Asn Glu Asp Gly Thr Trp Pro Arg Gly Pro		
210	215	220
Ser Thr Pro Asn Val Lys Asn Ser Asn Tyr Cys Leu Pro Ser Tyr Thr		
225	230	235
Ala Tyr Lys Asn Tyr Asp Tyr Ser Glu Pro Gly Arg Asn Asn Glu Gln		
245	250	255
Pro Gly Leu Cys Gly Leu Ser Asn Leu Gly Asn Thr Cys Phe Met Asn		
260	265	270
Ser Ala Ile Gln Cys Leu Ser Asn Thr Pro Pro Leu Thr Glu Tyr Phe		
275	280	285
Leu Asn Asp Lys Tyr Gln Glu Leu Asn Phe Asp Asn Pro Leu Gly		
290	295	300
Met Arg Gly Glu Ile Ala Lys Ser Tyr Ala Glu Leu Ile Lys Gln Met		
305	310	315
Trp Ser Gly Lys Phe Ser Tyr Val Thr Pro Arg Ala Phe Lys Thr Gln		
325	330	335
Val Gly Arg Phe Ala Pro Gln Phe Ser Gly Tyr Gln Gln Gln Asp Cys		
340	345	350
Gln Glu Leu Ala Phe Leu Leu Asp Gly Leu His Glu Asp Leu Asn		
355	360	365
Arg Ile Arg Lys Lys Pro Tyr Ile Gln Leu Lys Asp Ala Asp Gly Arg		
370	375	380
Pro Asp Lys Val Val Ala Glu Glu Ala Trp Glu Asn His Leu Lys Arg		
385	390	395
Asn Asp Ser Ile Ile Val Asp Ile Phe His Gly Leu Phe Lys Ser Thr		
405	410	415
Leu Val Cys Pro Glu Cys Ala Lys Ile Ser Val Thr Phe Asp Pro Phe		
420	425	430
Cys Tyr Leu Thr Leu Pro Leu Pro Met Lys Lys Glu Arg Thr Leu Glu		
435	440	445
Val Tyr Leu Val Arg Met Asp Pro Leu Thr Lys Pro Met Gln Tyr Lys		
450	455	460
Val Val Val Pro Lys Ile Gly Asn Ile Leu Asp Leu Cys Thr Ala Leu		
465	470	475
Ser Ala Leu Ser Gly Ile Pro Ala Asp Lys Met Ile Val Thr Asp Ile		
485	490	495
Tyr Asn His Arg Phe His Arg Ile Phe Ala Met Asp Glu Asn Leu Ser		
500	505	510
Ser Ile Met Glu Arg Asp Asp Ile Tyr Val Phe Glu Ile Asn Ile Asn		
515	520	525
Arg Thr Glu Asp Thr Glu His Val Ile Ile Pro Val Cys Leu Arg Glu		
530	535	540
Lys Phe Arg His Ser Ser Tyr Thr His His Thr Gly Ser Ser Leu Phe		
545	550	555
Gly Gln Pro Phe Leu Met Ala Val Pro Arg Asn Asn Thr Glu Asp Lys		
565	570	575
Leu Tyr Asn Leu Leu Leu Leu Arg Met Cys Arg Tyr Val Lys Ile Ser		
580	585	590
Thr Glu Thr Glu Glu Thr Glu Gly Ser Leu His Cys Cys Lys Asp Gln		
595	600	605

Asn Ile Asn Gly Asn Gly Pro Asn Gly Ile His Glu Glu Gly Ser Pro  
 610 615 620  
 Ser Glu Met Glu Thr Asp Glu Pro Asp Asp Glu Ser Ser Gln Asp Gln  
 625 630 635 640  
 Glu Leu Pro Ser Glu Asn Glu Asn Ser Gln Ser Glu Asp Ser Val Gly  
 645 650 655  
 Gly Asp Asn Asp Ser Glu Asn Gly Leu Cys Thr Glu Asp Thr Cys Lys  
 660 665 670  
 Gly Gln Leu Thr Gly His Lys Lys Arg Leu Phe Thr Phe Gln Phe Asn  
 675 680 685  
 Asn Leu Gly Asn Thr Asp Ile Asn Tyr Ile Lys Asp Asp Thr Arg His  
 690 695 700  
 Ile Arg Phe Asp Asp Arg Gln Leu Arg Leu Asp Glu Arg Ser Phe Leu  
 705 710 715 720  
 Ala Leu Asp Trp Asp Pro Asp Leu Lys Lys Arg Tyr Phe Asp Glu Asn  
 725 730 735  
 Ala Ala Glu Asp Phe Glu Lys His Glu Ser Val Glu Tyr Lys Pro Pro  
 740 745 750  
 Lys Lys Pro Phe Val Lys Leu Lys Asp Cys Ile Glu Leu Phe Thr Thr  
 755 760 765  
 Lys Glu Lys Leu Gly Ala Glu Asp Pro Trp Tyr Cys Pro Asn Cys Lys  
 770 775 780  
 Glu His Gln Gln Ala Thr Lys Lys Leu Asp Leu Trp Ser Leu Pro Pro  
 785 790 795 800  
 Val Leu Val Val His Leu Lys Arg Phe Ser Tyr Ser Arg Tyr Met Arg  
 805 810 815  
 Asp Lys Leu Asp Thr Leu Val Asp Phe Pro Ile Asn Asp Leu Asp Met  
 820 825 830  
 Ser Glu Phe Leu Ile Asn Pro Asn Ala Gly Pro Cys Arg Tyr Asn Leu  
 835 840 845  
 Ile Ala Val Ser Asn His Tyr Gly Gly Met Gly Gly Gly His Tyr Thr  
 850 855 860  
 Ala Phe Ala Lys Asn Lys Asp Asp Gly Lys Trp Tyr Tyr Phe Asp Asp  
 865 870 875 880  
 Ser Ser Val Ser Thr Ala Ser Glu Asp Gln Ile Val Ser Lys Ala Ala  
 885 890 895  
 Tyr Val Leu Phe Tyr Gln Arg Gln Asp Thr Phe Ser Gly Thr Gly Phe  
 900 905 910  
 Phe Pro Leu Asp Arg Glu Thr Lys Gly Ala Ser Ala Ala Thr Gly Ile  
 915 920 925  
 Pro Leu Glu Ser Asp Glu Asp Ser Asn Asp Asn Asp Asn Asp Ile Glu  
 930 935 940  
 Asn Glu Asn Cys Met His Thr Asn  
 945 950

<210> 502  
 <211> 4611  
 <212> DNA  
 <213> Homo sapiens

<400> 502  
 tggagaaga tggcggaagg cggagcggcg gatctggaca cccagcgggc tgacatcgcg 60  
 acgctgctca aaacctcgct ccggaagg gacacctggt acctagtcga tagtcgctgg 120  
 ttcaaacagt ggaaaaaata tgttggttt gacagttggg acaaatacca gatgggagat 180  
 caaaatgtgt atcctggacc cattgataac tctggacttc tcaaagatgg tgatgccag 240  
 tcaacttaagg aacatcttat tgaatgaattg gattacatac tgttgccaac tgaagggttg 300  
 aataaacttg tcagctggta cacattgatg gaagggtcaag agccaatagc acgaaagggtg 360  
 gttgaacagg gtatgtttgt aaagcactgc aaagtagaag tatatctcac agaattgaag 420  
 ctatgtgaaa atggaacat gaataatgtt gtaactcgaa gatttagcaa agctgacaca 480  
 atagatacaa ttgaaaagga aataagaaaa atcttcagta ttccagatga aaaggagacc 540

agattgtgga acaaatatcat gagtaacaca tttgaaccac tgaataaacc agacagcacc 600  
 attcaggatg ctgggtttata ccaaggacag gtattagtga tagaacagaa aaatgaagat 660  
 ggaacatggc caaggggtcc ttctactcct aatgtgaaaa actcaaatga ctgtcttcca 720  
 tcatataccg cttataagaa ctatgattat tcggaacctg gaagaaacaa tgaacagcca 780  
 ggctctgtg gcctaagtaa cttgggaaat acgtgtttca tgaactcagc tattcagtg 840  
 ttgagcaaca cacctccact tactgagtat ttccctcaatg ataagtatca agaagaactg 900  
 aattttgaca atcccttagg aatgagaggt gaaatagcta aatcttatgc cgaactgac 960  
 aagcaaatgt ggtctggaaa gtttagctac gtcaccccaa gagcctttaa gacacaggta 1020  
 ggacgttttg cacctcagtt ctctggatat cagcagcaag actgtcaaga actgttagct 1080  
 ttctatttag atggattaca tgaggatttg aatagaatta ggaaaaaacc atatatacaa 1140  
 ttaaagatg cagatggaag gccagataag gtggttgccg aagaagcctg ggaaaaccat 1200  
 ttaaaccgaa atgattctat catagtagat atatttcatg gccttttcaa atcaacttta 1260  
 gtttgtcctg agtgtgctaa gatttcagta acatttgcac ctttttgta ctgtacactt 1320  
 ccattgcccc tgaaaaaaga acgcaccttg gaagtttact tagttagaat ggatccgctt 1380  
 accaaacctt tgcagtacaa agtggttgct cccaaaattg gaaacatatt agatctttgt 1440  
 acagcattgt ctgctttgtc aggaatacct gcagataaga tgatagttac tgatatatac 1500  
 aatcatagat ttcacagaat attcgcctatg gatgaaaacc ttagtagtat tatggaacgg 1560  
 gatgatattt atgtgtttga aattaacatc aataggacag aagatacaga gcacgtgatt 1620  
 attcctgttt gcctaagaga aaaattcaga cactcgagtt ataccacca tactggttct 1680  
 tcaacttttg gtcagccctt tcttatggtc tgaccacgaa acaatactga agacaaactt 1740  
 tataatctcc tgctcttgag aatgtgccga tatgtcaaaa tatctactga aactgaagaa 1800  
 actgaaggat ccctacactg ctgtaaggac caaaatatta atgggaatgg cccaaatggc 1860  
 atacatgaag aaggctcacc aagtgaatg gaaacagatg agccagatga tgaatccagc 1920  
 caggatcaag aacttccctc agagaatgaa aacagtcagt ctgaagattc agttggagga 1980  
 gataatgatt ctgaaaatgg attatgtact gaggatactt gcaaaggta actcacggga 2040  
 cacaaaaaac gattgtttac attccagttc aacaacttag gcaatactga tatcaactac 2100  
 atcaaagatg ataccaggca tataagattt gatgataagg agcttaggct agatgaaaga 2160  
 tcttttcttg ctttggtatt ggatcctgat ttgaaaaaaa gatattttga tgaaaatgct 2220  
 gctgaggact ttgaaaaaca tgaaagtgtg gagtataaac ctctaaaaa accctttgtg 2280  
 aaattaaaag attgcattga actttttaca acaaaagaaa agctaggtgc tgaagatccc 2340  
 tggtattgtc cgaattgtaa agaacatcag caagccacaa agaaattgga tttatggtcc 2400  
 ctgctccag tactgtagt acatctcaag cgattttctt acagtcgata catgagagac 2460  
 aagttggata ccttagttga ttttctatc aatgacttgg atatgtcgga attcttaatt 2520  
 aatccaaatg caggtccttg ccgtatataat ctgattgctg tttccaacca ctatggaggg 2580  
 atgggaggag gacactatac tgcttttgca aaaaataaag atgatggaaa atggtactat 2640  
 tttgatgaca gtagtgtctc cactgcatct gaagaccaa ttgtgtccaa agcagcatat 2700  
 gtactcttct accagagaca agacacttct agtggaaact gctttttcc tcttgaccga 2760  
 gaaactaaa ggtgttcagc tgccactggc atcccattag aaagtgatga agatagcaat 2820  
 gataatgaca atgatataga aaatgaaaac tgtatgcaca ctaactaatg aaagtcctag 2880  
 aagccataaa agagacactt tctgtctggt ggtatctatg gaaatgatga agttaccac 2940  
 cacattaaaa caaaagtctg agatggggag tttcagataa ccgaatgtaa atctttatc 3000  
 agattttaac ttgtgcagta cttgaagtga aacacaatga aaactttaac agaaattgtc 3060  
 tcttaataca tttacagttc tgtatttaca agctaaatat atataggaaa tcacaaataa 3120  
 atccctttta agtttgctgc tgttttgatt aattatgttt cctttaattt tttggatgtg 3180  
 agttgatgac aaatgttaaa tttgtggatg ttggtcattt attttgctct aataatactg 3240  
 caagaaaact atggctgaac ttagtttttg gataaactag ttcatgtgca ttaggctgcc 3300  
 gtatatacfa ctatgatatt tagctgcagt atcaatgtgg cataaatact gcggcagcat 3360  
 tcttcgaaaa ccgcattttc aaatttatcc ccgtaatcaa tgtgggccta ttaaaaaacc 3420  
 aaatcatgat acaagaaaca accttggaag agttatttcc ctttgtaata cttttaaaaa 3480  
 tccagagtat tatattcatt ttactgccac tgtggaaaca ggttctagat ttcatactcc 3540  
 atctaaagtc atttttcagt tacatgtaag tattatttac tgctattggg atctattctt 3600  
 cagatattca aaaacatttt ttgcttttaa atgcatact ttaattgggt gttggtccaa 3660  
 aattaaaatt tttgctgtct gtttttctt accccatttt gtggaattt ggcaacttga 3720  
 gctctcccaa ttattgtaat ataaggacag acataatagt attctgtacc catagtaata 3780  
 attgcatcaa gcttagatga gaaattttt tcatatactg gcctttaaat cattaatgga 3840  
 caattggcta taaaggtagg tctgttaact ttctttgtgt gttcctgatg gaattcacca 3900  
 tagccttaca gcttttctca gaaggttaact tgttataaga gaaatggcat ttgcatgttc 3960  
 cagagtccag acctgtacag tatataaagc ataaacacct tgaatttgat tttattttgc 4020  
 cacattcaag gatccaggat gccaaaaata tgtgtgaaca tttgaaacat ttttattttgc 4080  
 tgctttttcc ctttgatata gttgaaagaa tgtattgtaa attggcatc aatactgcct 4140  
 ttaatagaaa acctaaatgc tggggcacat ttcacatata gactacctga gaaattgctt 4200

```

tgtgtccac  tgttattaaa  gcaaaaagta  taatgttctt  cacttgaact  aatcaaatac  4260
aatagattat  aaattgtgta  ttgtaaataa  aagttcactc  tgtgagtgca  catttttgta  4320
aattatttat  ttatgttagc  atttaaaagt  ttaaaaaaaa  tgcatttgac  caggataaat  4380
gaggtatggt  gatcatggct  ttgctttata  tcttgatatt  aaagctgggt  tatcatcctg  4440
gtatttttaa  agctgctttg  ggggtttttt  tttctttttt  ttttttttaa  tgtaactaa  4500
cctcctgcat  gacagaggtt  aaaattttgt  ctgcacttga  gttcacttga  gtttacattt  4560
gaaatgtgca  tgtttattta  tatagatttc  aataaagcta  ttcaaggcct  t  4611

```

&lt;210&gt; 503

&lt;211&gt; 187

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 503

```

Met Val Gly Ser Leu Asn Cys Ile Val Ala Val Ser Gln Asn Met Gly
 1          5          10          15
Ile Gly Lys Asn Gly Asp Leu Pro Trp Pro Pro Leu Arg Asn Glu Phe
          20          25          30
Arg Tyr Phe Gln Arg Met Thr Thr Thr Ser Ser Val Glu Gly Lys Gln
          35          40          45
Asn Leu Val Ile Met Gly Lys Lys Thr Trp Phe Ser Ile Pro Glu Lys
          50          55          60
Asn Arg Pro Leu Lys Gly Arg Ile Asn Leu Val Leu Ser Arg Glu Leu
          65          70          75          80
Lys Glu Pro Pro Gln Gly Ala His Phe Leu Ser Arg Ser Leu Asp Asp
          85          90          95
Ala Leu Lys Leu Thr Glu Gln Pro Glu Leu Ala Asn Lys Val Asp Met
          100          105          110
Val Trp Ile Val Gly Gly Ser Ser Val Tyr Lys Glu Ala Met Asn His
          115          120          125
Pro Gly His Leu Lys Leu Phe Val Thr Arg Ile Met Gln Asp Phe Glu
          130          135          140
Ser Asp Thr Phe Phe Pro Glu Ile Asp Leu Glu Lys Tyr Lys Leu Leu
          145          150          155          160
Pro Glu Tyr Pro Gly Val Leu Ser Asp Val Gln Glu Glu Lys Gly Ile
          165          170          175
Lys Tyr Lys Phe Glu Val Tyr Glu Lys Asn Asp
          180          185

```

&lt;210&gt; 504

&lt;211&gt; 3900

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 504

```

cgatggccct  gccagtcctc  agacagaacc  tactacgtgc  ggcggcagct  ggggcgggaa  60
ggcgggcgct  gggggcgctg  cggccgctgc  agcgcagggt  ccacctggct  ggctgcacct  120
gtggaggagg  aggtggattt  caggcttccc  gtgactgga  agaactggct  caaaaccgct  180
tgcctcgagc  gggctgagct  ggaggcagcg  aggccgcccg  acgcaggctt  ccggcgagac  240
atggcagggc  aaggatggca  gcccggcggc  agggcccggc  gaggagcgcg  aaccgcgggc  300
cgcagttccc  aggcgtctgc  gggcgcgagc  acgccgcgac  cctgcgtgcg  ccggggcggg  360
ggggcggggc  ctgcctgca  caaataggga  cgagggggcg  gggcgccac  aatttcgctc  420
caaaacttgac  cgcgcgttct  gctgtaacga  gcgggctcgg  aggtcctccc  gctgctgtca  480
tggttggttc  gctaaactgc  atcgtcgtct  tgtcccagaa  catgggcatc  ggcaagaacg  540
gggacctgac  ctggccaccg  ctgaggaatg  aattcagata  tttccagaga  atgaccacaa  600
cctcttcagt  agaaggtaaa  cagaatctgg  tgattatggg  taagaagacc  tggttctcca  660
ttcctgagaa  gaatcgacct  ttaaagggtg  gaattaattt  agttctcagc  agagaactca  720
aggaacctcc  acaaggagct  cattttcttt  ccagaagtct  agatgatgcc  ttaaaactta  780
ctgaacaacc  agaattagca  aataaagtag  acatggctct  gatagttggg  ggcagttctg  840
tttataagga  agccatgaat  caccaggcc  atcttaaact  atttgtgaca  aggatcatgc  900

```

```

aagactttga aagtgcacag ttttttccag aaattgattt ggagaaatat aaacttctgc 960
cagaataccc aggtgttctc tctgatgtcc aggaggagaa aggcattaag tacaaatttg 1020
aagtatatga gaagaatgat taatatgaag gtgttttcta gtttaagttg tccccctcc 1080
ctctgaaaaa agtatgtatt tttacattag aaaaggtttt ttgttgactt tagatctata 1140
attattttcta agcaacttgt ttttattccc cactactctt ctctctatca gataccattt 1200
atgagacatt ctgtctataa ctaagtgtct ctccaagacc ccaactgagt ccccgacc 1260
tgctacagtg agctgccatt ccacacccat cacatgtggc actcttgcca gtccttgaca 1320
ttgtcgggct tttcacatgt tggtaatatt tattaaagat gaagatccac atacccttca 1380
actgagcagt ttcactagtg gaaataccaa aagcttccta cgtgtatata cagaggtttg 1440
tagataaatg ttgccacctt gtttgaaca gtgaaaaatt gaaaacaacc tgggaagcca 1500
gtgatgggaa aatgagtatg tttctgtctt agattgggga acccaaagca gattgcaaga 1560
ctgaaatttc agtgaaagca gtgtatttgc taggtcatac cagaaatcat caattgaggt 1620
acggagaaac tgaactgaga aggttaagaaa agcaatttaa agtcagcgag caggttctca 1680
ttgataacaa gctccatact gctgagatac agggaaatgg aggggggaaa gctggagtat 1740
tgatcccgcc cccctccttg gttgtcagct ccctgtcctg tgtgtgggag gaacatagtc 1800
cagctgtctc atagcaagtc tcaggtgttt gcagtaagaa gctgctggca tgcacgggaa 1860
cagtgaatgc caaacactta aagcaattcg atgtttaagt atgtaagttc tttttttttt 1920
agacagcggt tcgctcttgt tgcccaggct agcatgcaat ggtgtgacct cggcttactg 1980
caacctccgc ctcccgatg tcaagcgatt ccctgcctc aggtcccaa gtactagga 2040
ccaggtgcgc gccaccacgc ccggctaatt tttgtatttt gtatttttag tagagatggg 2100
gtttcaccat gttggtcagg ctagtctcga actcgtgacc gcaagcgatt caccacctc 2160
agcctcccaa agtgctggga ttaccggctt gagccaccac acccggcaca tcttcattct 2220
ttttatgtag taaaaagtat aaggccacac atggtttatt tgaagtattt tataatttaa 2280
aaaaatacac aagcaggaaa accaattata agttcaagtg agggatgatg gttgcttgaa 2340
ccaaaggggt gcatgtagta agaaattgtg atttaagata tattttaag ttataagtag 2400
caggatattc tgatggagtt tgactttggt tttgggcca gggagtttca gatgccttg 2460
agaaatgaat gaagtagaga gaaaataaaa gaaaaaccag ccaggcacag tggctcacag 2520
ctgtaatccc agcgctttgg gaggctaagg caggcagatc acttgagacc agcttgggca 2580
acatggcaaa gcccctctc tacaaaaaac acaaaaatta gctgggcatt gtggcgaca 2640
cctgtattcc catctagtca ggaagctgag atggaagaat taattgagcc cactgagtc 2700
aggctgcagt gactcgtgat tgtgccctg cactccagcc ggggtgacag aagagacctt 2760
gtctcgaaaa cgaatctgaa aacaattgaa ccattgcctc ataattctag aaagtattt 2820
tcaactgata aatctatatt caccacaata atcaagggtg aaggtaaaat aatacathtt 2880
tagacaagca aagactcagg ggttacctcc atgtgccctt tttagggaag ctgttgagga 2940
aaatactcca gcaaaatgaa ggagtacaca aaccagagaa tgacatgaat ccagcaaata 3000
ggatccaaca caggcaatat tccagctatg gagctagctt taaaaaggaa cagtaaaaat 3060
attaatcggt tagctgggtg gaatggccca tgctgtagt cccagctact caggaggctc 3120
agcagcagga cgacttgagc ccaagagttc cagaccagcc tggccacctt agtgagatcc 3180
cttctcttaa aaataataac ttattgccag atttggggca tttggaaaga agttcattga 3240
agataaagca aaagtaaaaa aaaaaaaaaa aaaaacaagg ggaaaggggt ggttaggcaa 3300
tcattctagg gcagaaagaa gtacaggata ggaagagcat aatacactgt ttttctcaac 3360
aaggagcagt atgtacacag tcataatgat gtgactgctt agcccctaaa tatggtaact 3420
actctgggag aatatgggag gaaaagtga gatttgtgat gtgtaagagc taatcctcat 3480
ctgtcatatc cagaaatcac tatataatat ataataatga aatgactaag ttatgtgagg 3540
aaaaaacag aagacattgc taaaagagtt aaaagtcatt gctctggaga attaggaggg 3600
atggggcagg ggactgttag gatgcattat aaactgaaaa gcctttttta aattttatgt 3660
attaatatat gcattcactt gaaaaactaa aaaaaacaa taatttgga aaacccatga 3720
aggtaactaa cggaaggaaa aactaagaga atgaaaagta tttgcctctg gaaagaacaa 3780
ctggcaggac tgttgttttc attgtaagac ttttgagacc atttaattgt acttaaccat 3840
ttcatctat ttctttaata agaacaattc catcttaata aagagttaca cttgttaata 3900

```

<210> 505  
 <211> 1616  
 <212> PRT  
 <213> Homo sapiens

<400> 505  
 Met Pro Ala Arg Thr Ala Pro Ala Arg Val Pro Thr Leu Ala Val Pro  
 1 5 10 15  
 Ala Ile Ser Leu Pro Asp Asp Val Arg Arg Arg Leu Lys Asp Leu Glu

-266-

Val Glu Phe Leu Gln Ser Asn Ser Asp Ser Thr Tyr Glu Asp Leu Ile  
 515 520 525  
 Asn Lys Ile Glu Thr Thr Val Pro Pro Ser Gly Leu Asn Leu Asn Arg  
 530 535 540  
 Phe Thr Glu Asp Ser Leu Leu Arg His Ala Gln Phe Val Val Glu Gln  
 545 550 555 560  
 Val Glu Ser Tyr Asp Glu Ala Gly Asp Ser Asp Glu Gln Pro Ile Phe  
 565 570 575  
 Leu Thr Pro Cys Met Arg Asp Leu Ile Lys Leu Ala Gly Val Thr Leu  
 580 585 590  
 Gly Gln Arg Arg Ala Gln Ala Arg Arg Gln Thr Ile Arg His Ser Thr  
 595 600 605  
 Arg Glu Lys Asp Arg Gly Pro Thr Lys Ala Thr Thr Thr Lys Leu Val  
 610 615 620  
 Tyr Gln Ile Phe Asp Thr Phe Phe Ala Glu Gln Ile Glu Lys Asp Asp  
 625 630 635 640  
 Arg Glu Asp Lys Glu Asn Ala Phe Lys Arg Arg Arg Cys Gly Val Cys  
 645 650 655  
 Glu Val Cys Gln Gln Pro Glu Cys Gly Lys Cys Lys Ala Cys Lys Asp  
 660 665 670  
 Met Val Lys Phe Gly Gly Ser Gly Arg Ser Lys Gln Ala Cys Gln Glu  
 675 680 685  
 Arg Arg Cys Pro Asn Met Ala Met Lys Glu Ala Asp Asp Asp Glu Glu  
 690 695 700  
 Val Asp Asp Asn Ile Pro Glu Met Pro Ser Pro Lys Lys Met His Gln  
 705 710 715 720  
 Gly Lys Lys Lys Lys Gln Asn Lys Asn Arg Ile Ser Trp Val Gly Glu  
 725 730 735  
 Ala Val Lys Thr Asp Gly Lys Lys Ser Tyr Tyr Lys Lys Val Cys Ile  
 740 745 750  
 Asp Ala Glu Thr Leu Glu Val Gly Asp Cys Val Ser Val Ile Pro Asp  
 755 760 765  
 Asp Ser Ser Lys Pro Leu Tyr Leu Ala Arg Val Thr Ala Leu Trp Glu  
 770 775 780  
 Asp Ser Ser Asn Gly Gln Met Phe His Ala His Trp Phe Cys Ala Gly  
 785 790 795 800  
 Thr Asp Thr Val Leu Gly Ala Thr Ser Asp Pro Leu Glu Leu Phe Leu  
 805 810 815  
 Val Asp Glu Cys Glu Asp Met Gln Leu Ser Tyr Ile His Ser Lys Val  
 820 825 830  
 Lys Val Ile Tyr Lys Ala Pro Ser Glu Asn Trp Ala Met Glu Gly Gly  
 835 840 845  
 Met Asp Pro Glu Ser Leu Leu Glu Gly Asp Asp Gly Lys Thr Tyr Phe  
 850 855 860  
 Tyr Gln Leu Trp Tyr Asp Gln Asp Tyr Ala Arg Phe Glu Ser Pro Pro  
 865 870 875 880  
 Lys Thr Gln Pro Thr Glu Asp Asn Lys Phe Lys Phe Cys Val Ser Cys  
 885 890 895  
 Ala Arg Leu Ala Glu Met Arg Gln Lys Glu Ile Pro Arg Val Leu Glu  
 900 905 910  
 Gln Leu Glu Asp Leu Asp Ser Arg Val Leu Tyr Tyr Ser Ala Thr Lys  
 915 920 925  
 Asn Gly Ile Leu Tyr Arg Val Gly Asp Gly Val Tyr Leu Pro Pro Glu  
 930 935 940  
 Ala Phe Thr Phe Asn Ile Lys Leu Ser Ser Pro Val Lys Arg Pro Arg  
 945 950 955 960  
 Lys Glu Pro Val Asp Glu Asp Leu Tyr Pro Glu His Tyr Arg Lys Tyr  
 965 970 975  
 Ser Asp Tyr Ile Lys Gly Ser Asn Leu Asp Ala Pro Glu Pro Tyr Arg  
 980 985 990  
 Ile Gly Arg Ile Lys Glu Ile Phe Cys Pro Lys Lys Ser Asn Gly Arg



995	1000	1005
Pro Asn Glu Thr Asp Ile Lys Ile Arg Val Asn Lys Phe Tyr Arg Pro		
1010	1015	1020
Glu Asn Thr His Lys Ser Thr Pro Ala Ser Tyr His Ala Asp Ile Asn		
1025	1030	1035
Leu Leu Tyr Trp Ser Asp Glu Glu Ala Val Val Asp Phe Lys Ala Val		1040
	1045	1050
Gln Gly Arg Cys Thr Val Glu Tyr Gly Glu Asp Leu Pro Glu Cys Val		1055
	1060	1065
Gln Val Tyr Ser Met Gly Gly Pro Asn Arg Phe Tyr Phe Leu Glu Ala		1070
	1075	1080
Tyr Asn Ala Lys Ser Lys Ser Phe Glu Asp Pro Pro Asn His Ala Arg		1085
	1090	1095
Ser Pro Gly Asn Lys Gly Lys Gly Lys Gly Lys Gly Lys Pro		1100
1105	1110	1115
Lys Ser Gln Ala Cys Glu Pro Ser Glu Pro Glu Ile Glu Ile Lys Leu		1120
	1125	1130
Pro Lys Leu Arg Thr Leu Asp Val Phe Ser Gly Cys Gly Gly Leu Ser		1135
	1140	1145
Glu Gly Phe His Gln Ala Gly Ile Ser Asp Thr Leu Trp Ala Ile Glu		1150
	1155	1160
Met Trp Asp Pro Ala Ala Gln Ala Phe Arg Leu Asn Asn Pro Gly Ser		1165
	1170	1175
Thr Val Phe Thr Glu Asp Cys Asn Ile Leu Leu Lys Leu Val Met Ala		1180
1185	1190	1195
Gly Glu Thr Thr Asn Ser Arg Gly Gln Arg Leu Pro Gln Lys Gly Asp		1200
	1205	1210
Val Glu Met Leu Cys Gly Gly Pro Pro Cys Gln Gly Phe Ser Gly Met		1215
	1220	1225
Asn Arg Phe Asn Ser Arg Thr Tyr Ser Lys Phe Lys Asn Ser Leu Val		1230
	1235	1240
Val Ser Phe Leu Ser Tyr Cys Asp Tyr Tyr Arg Pro Arg Phe Phe Leu		1245
	1250	1255
Leu Glu Asn Val Arg Asn Phe Val Ser Phe Lys Arg Ser Met Val Leu		1260
1265	1270	1275
Lys Leu Thr Leu Arg Cys Leu Val Arg Met Gly Tyr Gln Cys Thr Phe		1280
	1285	1290
Gly Val Leu Gln Ala Gly Gln Tyr Gly Val Ala Gln Thr Arg Arg Arg		1295
	1300	1305
Ala Ile Ile Leu Ala Ala Ala Pro Gly Glu Lys Leu Pro Leu Phe Pro		1310
	1315	1320
Glu Pro Leu His Val Phe Ala Pro Arg Ala Cys Gln Leu Ser Val Val		1325
	1330	1335
Val Asp Asp Lys Lys Phe Val Ser Asn Ile Thr Arg Leu Ser Ser Gly		1340
1345	1350	1355
Pro Phe Arg Thr Ile Thr Val Arg Asp Thr Met Ser Asp Leu Pro Glu		1360
	1365	1370
Val Arg Asn Gly Ala Ser Ala Leu Glu Ile Ser Tyr Asn Gly Glu Pro		1375
	1380	1385
Gln Ser Trp Phe Gln Arg Gln Leu Arg Gly Ala Gln Tyr Gln Pro Ile		1390
	1395	1400
Leu Arg Asp His Ile Cys Lys Asp Met Ser Ala Leu Val Ala Ala Arg		1405
	1410	1415
Met Arg His Ile Pro Leu Ala Pro Gly Ser Asp Trp Arg Asp Leu Pro		1420
1425	1430	1435
Asn Ile Glu Val Arg Leu Ser Asp Gly Thr Met Ala Arg Lys Leu Arg		1440
	1445	1450
Tyr Thr His His Asp Arg Lys Asn Gly Arg Ser Ser Ser Gly Ala Leu		1455
	1460	1465
Arg Gly Val Cys Ser Cys Val Glu Ala Gly Lys Ala Cys Asp Pro Ala		1470
	1475	1480
		1485

Ala Arg Gln Phe Asn Thr Leu Ile Pro Trp Cys Leu Pro His Thr Gly  
 1490 1495 1500  
 Asn Arg His Asn His Trp Ala Gly Leu Tyr Gly Arg Leu Glu Trp Asp  
 1505 1510 1515 1520  
 Gly Phe Phe Ser Thr Thr Val Thr Asn Pro Glu Pro Met Gly Lys Gln  
 1525 1530 1535  
 Gly Arg Val Leu His Pro Glu Gln His Arg Val Val Ser Val Arg Glu  
 1540 1545 1550  
 Cys Ala Arg Ser Gln Gly Phe Pro Asp Thr Tyr Arg Leu Phe Gly Asn  
 1555 1560 1565  
 Ile Leu Asp Lys His Arg Gln Val Gly Asn Ala Val Pro Pro Pro Leu  
 1570 1575 1580  
 Ala Lys Ala Ile Gly Leu Glu Ile Lys Leu Cys Met Leu Ala Lys Ala  
 1585 1590 1595 1600  
 Arg Glu Ser Ala Ser Ala Lys Ile Lys Glu Glu Glu Ala Ala Lys Asp  
 1605 1610 1615

<210> 506  
 <211> 5434  
 <212> DNA  
 <213> Homo sapiens

<400> 506

```

cgccccgctg ggggggggtg gtccccgcct tgcgcacgct tgttccctgg gcatggccgg 60
ctccgttcca tccttctgca cagggtatcg cctctctccg tttggtacat cccctcctcc 120
cccacgcccg gactgggggtg gtagacgcgc ctccgctcat cgccccctcc catcggtttc 180
cgcgcgaaaa gccgggggccc ctgcgctgcc gccgcgcgct ctgctgaagc ctccgagatg 240
ccggcgcgta ccgccccagc ccgggtgccc aactggccg tcccgccat ctgctgccc 300
gacgatgtcc gcaggcgggt caaagatttg gaaagagaca gcttaacaga aaaggaatgt 360
gtgaaggaga aattgaatct cttgcacgaa gaattatccg aggagggcta cctggctaaa 420
tgtgacttgg aaaccaaatt acgtaaagaa tttctgcaaa cagaaataaa gaatcagtta 480
gtcaaatccc ttttaataaa agatttgtcc ttggagaacg gtgctcatgc ttacaaccgg 540
gaagtgaatg gacgtctaga aaacgggaac caagcaagaa gtgaagcccg tagagtggga 600
atggcagatg ccaacagccc ccccaaacc ctttccaaac ctgcacgcc caggaggagc 660
aagtccgatg gagaggctaa gcctgaacct tcacctagcc ccaggattac aaggaaaagc 720
accagtcaca ccaccatcac atctcatttt gcaaagggcc ctgccaaacg gaaacctcag 780
gaagagtctg aaagagccaa atcggatgag tccatcaagg aagaagacaa agaccagatg 840
gagaagagac gtagagttac atccagagaa cgagttgcta gaccgcttcc tgcagaagaa 900
cctgaaagag caaaatcagg aacgcgcact gaaaaggaag aagaaagaga tgaaaaagaa 960
gaaaagagac tccgaagtca aaccaaagaa ccaacacca aacagaaact gaaggaggag 1020
ccggacagag aagccagggc aggcgtgcag gctgacgagg acgaagatgg agacgagaaa 1080
gatgagaaga agcacagaag tcaacccaaa gatctagctg ccaaaccggag gcccgaaagaa 1140
aaagaacctg aaaaagttaa tccacagatt tctgatgaaa aagacgagga tgaaaaggag 1200
gagaagagac gcaaaacgac ccccaaagaa ccaacggaga aaaaaatggc tcgcgcaaaa 1260
acagtcatga actccaagac ccacctccc aagtgcattc agtgcgggca gtacctggac 1320
gacctgacc tcaaatatgg gcagaccca ccagacgagg tggatgagcc acagatgctg 1380
acaaatgaga agctgtccat ctttgatgcc aacgagtctg gctttgagag ttatgaggcg 1440
cttccccagc acaaactgac ctgcttcagt gtgtactgta agcacggtca cctgtgtccc 1500
atcgacaccg gcctcatcga gaagaatct gaactcttct tttctggttc agcaaaaacca 1560
atctatgatg atgaccgctc tcttgaaggt ggtgttaatg gcaaaaatct tggccccata 1620
aatgaatggg ggatcactgg ctttgatgga ggtgaaaagg cctcatcgg cttcagcacc 1680
tcatttgccg aatacattct gatggatccc agtcccagat atgcgcccat atttgggctg 1740
atgcaggaga agatctacat cagcaagatt gtggtggagt tcctgcagag caattccgac 1800
tcgacctatg aggacctgat caacaagatc gagaccacgg ttctctcttc tggcctcaac 1860
ttgaaccgct tcacagagga ctccctcctg cgacacgcgc agtttgtggg ggagcagggtg 1920
gagagttatg acgaggccgg ggacagtgat gagcagccca tcttctctgac gccctgcatg 1980
cgggacctga tcaagctggc tggggtcacg ctgggacaga ggcgagccca ggcgaggcgg 2040
cagaccatca ggcattctac cagggagaag gacaggggac ccacgaaagc caccaccacc 2100
aagctggtct accagatctt cgatactttc ttcgcagagc aaattgaaaa ggatgacaga 2160
gaagacaagg agaacgcctt taagcgccgg cgatgtggcg tctgtgaggt gtgtcagcag 2220

```

```

cctgagtgtg ggaaatgtaa agcctgcaag gacatgggta aatttggtgg cagtggacgg 2280
agcaagcagg cttgccaaaga gcgagggtgt cccaatatgg ccatgaagga ggcagatgac 2340
gatgaggaag tcgatgataa catcccagag atgccgtcac ccaaaaaaat gcaccagggg 2400
aagaagaaga aacagaacaa gaatcgcatc tcttgggtcg gagaagccgt caagactgat 2460
gggaagaaga gttactataa gaaggtgtgc attgatgcgg aaaccctgga agtggggggac 2520
tgtgtctctg ttattccaga tgattcctca aaaccgctgt atctagcaag ggtcacggcg 2580
ctgtgggagg acagcagcaa cgggcagatg ttccacgccc actggttctg cgctgggaca 2640
gacacagtcc tcggggccac gtcggaccct ctggagctgt tcttggtgga tgaatgtgag 2700
gacatgcagc ttcatatat ccacagcaaa gtgaaagtca tctacaaagc cccctccgaa 2760
aactgggcca tggaggagg catggatccc gagtccctgc tggaggggga cgacgggaag 2820
acctacttct accagctgtg gtatgatcaa gactacgcga gattcgagtc cctccaaaa 2880
accagccaa cagaggacaa caagttcaaa ttctgtgtga gctgtgcccg tctggctgag 2940
atgaggcaaa aagaaatccc cagggtcctg gagcagctcg aggacctgga tagccgggtc 3000
ctctactact cagccaccaa gaacggcatc ctgtaccgag ttggtgatgg tgtgtacctg 3060
ccccctgagg ccttcagctt caacatcaag ctgtccagtc ccgtgaaacg cccacggaag 3120
gagcccgagg atgaggacct gtaccagag cactaccgga aatactccga ctacatcaaa 3180
ggcagcaacc tggatgcccc tgagccctac cgaattggcc ggatcaaaga gatcttctgt 3240
cccaagaaga gcaacggcag gcccaatgag actgacatca aaatccgggt caacaagttc 3300
tacaggcctg agaacaccca caagtccact ccagcgagct accacgcaga catcaacctg 3360
ctctactgga gcgacgagga ggccgtggtg gacttcaagg ctgtgcaggg ccgtgcacc 3420
gtggagtatg gggaggacct gcccgagtgc gtccaggtgt actccatggg cgccccaac 3480
cgcttctact ccttcagggc ctataatgca aagagcaaaa gctttgaaga tcctcccaac 3540
catgcccgta gccctggaaa caaagggaag ggcaaggga aagggaaggg caagcccaag 3600
tcccaagcct gtgagccgag cgagccagag atagagatca agctgcccac gctgcggacc 3660
ctggatgtgt tttctggctg cgggggggtt tcggagggat tccaccaagc aggcattctt 3720
gacacgctgt gggccatcga gatgtgggac cctgcggccc aggcgttccg gctgaacaac 3780
cccggctcca cagtgttcac agaggactgc aacatcctgc tgaagctggt catggctggg 3840
gagaccacca actcccgagg ccagcggctg cccagaagg gagacgtgga gatgctgtgc 3900
ggcgggcccg cctgccagg cttcagcggc atgaaccgct tcaattcgcg cactactcc 3960
aagttcaaaa actctctggt ggtttccttc ctacgtact gcgactacta ccggccccgg 4020
ttcttctctc tggagaatgt caggaaactt gtctccttca agcgtcccat ggtcctgaag 4080
ctcaccctcc gctgcctggt ccgcatgggc tatcagtga ccttcggcgt gctgcaggcc 4140
ggtcagtagc gcgtggccca gactaggagg cgggccatca tctggccgc ggccccctgga 4200
gagaagctcc ctctgttccc ggagccactg cacgtgtttg ctccccgggc ctgccagctg 4260
agcgtggtgg tggatgacaa gaagtttgtg agcaacataa ccagggtgag ctcggtctct 4320
ttccggacca tcacgggtgc agacacgatg tccgacctgc cggaggtgcg gaatggagcc 4380
tcggcactgg agatctccta caacggggag cctcagtcct ggttccagag gcagctccgg 4440
ggcgcacagt accagcccat cctcagggac cacatctgta aggacatgag tgcattggtg 4500
gctgcccga tcgggcacat ccccttggcc ccagggtcag actggcgca tctgcccac 4560
atcgagggtg ggctctcaga cggcaccatg gccagggaag tcgggtatac ccaccatgac 4620
aggaagaacg gccgcagcag ctctggggcc ctccgtgggg tctgctcctg cgtggaagcc 4680
ggcaaaagct gcgaccccg agccaggcag ttcaacacc tcatccccctg gtgctgccc 4740
cacaccggga accggcacia ccactgggct ggctctatg gaaggctcga gtgggacggc 4800
ttcttcagca caaccgtcac caaccggag cccatgggca agcagggccg cgtgctccac 4860
ccagagcagc accgtgtggt gagcgtgcgg gagtgtgccc gctcccaggg ctccctgac 4920
acctaccggc tcttcggcaa catcctggac aagcaccggc aggtgggcaa tgcggtgcca 4980
ccgcccctgg ccaaagccat tggcttgagg atcaagcttt gtatgttggc caaagcccga 5040
gagagtgcct cagctaaaaa aaaggaggag gaagctgcta aggactagt ctgccctccc 5100
gtcaccctgt tttctggcac caggaatccc caacatgcac tgatgttgtg tttttaacat 5160
gtcaatctgt ccgttcacat gtgtgttaca tgggtttgt ggcttggct gacatgaagc 5220
tgttgtgtga ggttcgctta tcaactaatg atttagtgat caaattgtgc agtactttgt 5280
gcattctgga ttttaaaagt tttttattat gcattatata aaatctacca ctgtatgagt 5340
ggaaattaag actttatgta gtttttatat gttgtaatat ttcttcaaat aaatctctcc 5400
tataaaccaa aaaaaaaaaa aaaaaaaaaa aaaa 5434

```

<210> 507  
<211> 107  
<212> PRT  
<213> Homo sapiens

<400> 507

Met Ala Arg Ala Ala Leu Ser Ala Ala Pro Ser Asn Pro Arg Leu Leu  
 1 5 10 15  
 Arg Val Ala Leu Leu Leu Leu Leu Val Ala Ala Gly Arg Arg Ala  
 20 25 30  
 Ala Gly Ala Ser Val Ala Thr Glu Leu Arg Cys Gln Cys Leu Gln Thr  
 35 40 45  
 Leu Gln Gly Ile His Pro Lys Asn Ile Gln Ser Val Asn Val Lys Ser  
 50 55 60  
 Pro Gly Pro His Cys Ala Gln Thr Glu Val Ile Ala Thr Leu Lys Asn  
 65 70 75 80  
 Gly Arg Lys Ala Cys Leu Asn Pro Ala Ser Pro Ile Val Lys Lys Ile  
 85 90 95  
 Ile Glu Lys Met Leu Asn Ser Asp Lys Ser Asn  
 100 105

<210> 508  
 <211> 1103  
 <212> DNA  
 <213> Homo sapiens

<400> 508

cacagagccc gggccgcagg cacctcctcg ccagctcttc cgctcctctc acagccgccca 60  
 gaccgcctg ctgagcccca tggccgcgc tgcctctctc gccgccccca gcaatccccg 120  
 gctcctgaga gtggcactgc tgctcctgct cctggtagcc gctggccggc gcgcagcagg 180  
 agcgtccgtg gccactgaac tgcgctgcca gtgcttgagc accctgcagg gaattcacc 240  
 caagaacatc caaagtgtga acgtgaagtc ccccgaccc cactgcgccc aaaccgaagt 300  
 catagccaca ctcaagaatg ggcggaaagc ttgcctcaat cctgcatccc ccatagttaa 360  
 gaaaatcatc gaaaagatgc tgaacagtga caaatccaac tgaccagaag ggaggaggaa 420  
 gctcactggt ggctgttctt gaaggaggcc ctgcccttat aggaacagaa gaggaaagag 480  
 agacacagct gcagaggcca cctggattgt gcctaagtgt tttgagcatc gcttaggaga 540  
 agtcttctat ttatttattt attcattagt ttgaagatt ctatgttaat attttagggt 600  
 taaaataatt aagggtatga ttaactctac ctgcacactg tcctattata ttcattcttt 660  
 ttgaaatgtc aaccccaagt tagttcaatc tggattcata tttaatttga aggtagaatg 720  
 ttttcaaattg ttctccagtc attatgttaa ttttctgag gagcctgcaa catgccagcc 780  
 actgtgatag aggtcggcgg atccaagcaa atggccaatg agatcattgt gaaggcaggg 840  
 gaatgtatgt gcacatctgt tttgtaactg tttagatgaa tgtcagttgt tattttattga 900  
 aatgatttca cagtgtgtgg tcaacatttc tcatgttgaa actttaagaa ctaaaatggt 960  
 ctaaaatatcc cttggacatt ttatgtcttt cttgtaaggc atactgcctt gtttaatggt 1020  
 agttttacag tgtttctggc ttagaacaata ggggcttaat tattgatgtt ttcatagaga 1080  
 atataaaaat aaagcactta tag 1103

<210> 509  
 <211> 107  
 <212> PRT  
 <213> Homo sapiens

<400> 509

Met Ala Arg Ala Thr Leu Ser Ala Ala Pro Ser Asn Pro Arg Leu Leu  
 1 5 10 15  
 Arg Val Ala Leu Leu Leu Leu Leu Val Ala Ala Ser Arg Arg Ala  
 20 25 30  
 Ala Gly Ala Pro Leu Ala Thr Glu Leu Arg Cys Gln Cys Leu Gln Thr  
 35 40 45  
 Leu Gln Gly Ile His Leu Lys Asn Ile Gln Ser Val Lys Val Lys Ser  
 50 55 60  
 Pro Gly Pro His Cys Ala Gln Thr Glu Val Ile Ala Thr Leu Lys Asn  
 65 70 75 80  
 Gly Gln Lys Ala Cys Leu Asn Pro Ala Ser Pro Met Val Lys Lys Ile  
 85 90 95  
 Ile Glu Lys Met Leu Lys Asn Gly Lys Ser Asn

100

105

<210> 510  
 <211> 1110  
 <212> DNA  
 <213> Homo sapiens

<400> 510  
 gacagagccc gggccacgga gctccttgcc agctctcctc ctgcacacgc cgctcgaacc 60  
 gcctgctgag ccccatggcc cgcgccacgc tctccgccgc cccagcaat ccccggtcc 120  
 tgcgggtggc gctgctgctc ctgctcctgg tggccgccag ccggcgcgca gcaggagcgc 180  
 ccctggccac tgaactgcgc tgccagtgtc tgcagaccct gcagggaatt cacctcaaga 240  
 acatccaaag tgtgaagggtg aagtcctccg gacccactg cgcctcaacc gaagtcatag 300  
 ccacactcaa gaatgggcag aaagcttgct tcaacccgc atcgcccatg gttaagaaaa 360  
 tcatcgaaaa gatgctgaaa aatggcaaat ccaactgacc agaaggagg aggaagctta 420  
 ttggtggctg ttctgaagg aggcctgccc ttacaggaac agaagaggaa agagagacac 480  
 agctgcagag gccacctggc ttgcgcctaa tgtgtttgag catacttagg agaagtcctc 540  
 tattttattt ttattttatt tttttgtttg ttttagaaga ttctatgtta atattttatg 600  
 tgtaaaataa ggttatgatt gaatctactt gcacactctc ccattatatt tattgtttat 660  
 tttaggtcaa acccaagtta gttcaatcct gattcatatt taatttgaag atagaagggt 720  
 tgcagatatt ctctagtcct ttgttaatat ttcttcgtga tgacatatca catgtcagcg 780  
 actgtgatag aggtcgagga atccaagaaa atggccagta agatcaatgt gacggcaggg 840  
 aaatgtatgt gtgtctatct tgtaactgta aagatgaatg tcagttgtta tttattgaaa 900  
 tgatttcaca gtgtgtggtc aacatttctc atgttgaaagc tttaagaact aaaatgttct 960  
 aaatatccct tggcatttta tgtctttctt gtaagatact gccttgttta atgttaatta 1020  
 tgcagtgttt ccctctgtgt tagagcagag aggtttcgat atttattgat gttttcacia 1080  
 agaacaggaa aataaaatat ttaaaaatat 1110

<210> 511  
 <211> 99  
 <212> PRT  
 <213> Homo sapiens

<400> 511  
 Met Thr Ser Lys Leu Ala Val Ala Leu Leu Ala Ala Phe Leu Ile Ser  
 1 5 10 15  
 Ala Ala Leu Cys Glu Gly Ala Val Leu Pro Arg Ser Ala Lys Glu Leu  
 20 25 30  
 Arg Cys Gln Cys Ile Lys Thr Tyr Ser Lys Pro Phe His Pro Lys Phe  
 35 40 45  
 Ile Lys Glu Leu Arg Val Ile Glu Ser Gly Pro His Cys Ala Asn Thr  
 50 55 60  
 Glu Ile Ile Val Lys Leu Ser Asp Gly Arg Glu Leu Cys Leu Asp Pro  
 65 70 75 80  
 Lys Glu Asn Trp Val Gln Arg Val Val Glu Lys Phe Leu Lys Arg Ala  
 85 90 95  
 Glu Asn Ser

<210> 512  
 <211> 1666  
 <212> DNA  
 <213> Homo sapiens

<400> 512  
 ctccataagg cacaaacttt cagagacacg agagcacaca agcttctagg acaagagcca 60  
 ggaagaaacc accggaagga accatctcac tgtgtgtaaa catgacttcc aagctggccg 120  
 tggctctctt ggcagccttc ctgatttctg cagctctgtg tgaagggtgca gttttgccaa 180  
 ggagtgtctaa agaacttaga tgtcagtgtc taaagacata ctccaaacct ttccacccca 240

```

aatttatcaa agaactgaga gtgattgaga gtggaccaca ctgcgccaac acagaaatta 300
ttgtaaagct tttgatgga agagagctct gtctggacc caaggaaaaac tgggtgcaga 360
gggttggtga gaagtttttg aagagggctg agaattcata aaaaaattca ttctctgtgg 420
tatccaagaa tcagtgaaga tgccagtga acttcaagca aatctacttc aacacttcat 480
gtattgtgtg ggtctgttgt agggttgcca gatgcaatac aagattcctg gttaaatttg 540
aatttcagta aacaatgaat agtttttcat tgtaccatga aatatccaga acatacttat 600
atgtaaagta ttattttatt gaatctacaa aaaacaacaa ataattttta aatataagga 660
ttttcctaga tattgcacgg gagaatatac aaatagcaaa attgaggcca agggccaaga 720
gaatatccga actttaattt caggaattga atgggtttgc tagaatgtga ttttggaagc 780
atcacataaa aatgatggga caataaattt tgccataaag tcaaatttag ctggaaatcc 840
tggatttttt tctgttaaat ctggcaaccc tagtctgcta gccaggatcc acaagtcctt 900
gttccactgt gccttggttt ctcctttatt tctaagtga aaaagtatta gccaccatct 960
tacctcacag tgatgttgtg aggacatgtg gaagcacttt aagttttttc atcataacat 1020
aaattatttt caagtgtaac ttattaacct atttattatt tatgtattta ttttaagcatc 1080
aaatatttgt gcaagaattt ggaaaaatag aagatgaatc attgattgaa tagttataaa 1140
gatgttatag taaattttatt ttattttaga tattaaatga tgttttatta gataaatttc 1200
aatcagggtt tttagattaa acaaacaac aattgggtac ccagttaaat tttcatttca 1260
gataaacaac aaataatttt ttagtataag tacattattg tttatctgaa attttaattg 1320
aactaacaat cctagtttga tactcccgat ctgtcattg ccagctgtgt tggtagtgt 1380
gtgttgaatt acggaataat gagttagaac tattaaaaca gccaaaactc cacagtcaat 1440
attagtaatt tcttgctggt tgaacttgt ttattatgta caaatagatt cttataatat 1500
tatttaaatg actgcatttt taaatacaag gctttatatt ttttaacttta agatgttttt 1560
atgtgctctc caaatttttt ttactgtttc tgattgtatg gaaatataaa agtaaatatg 1620
aaacatttaa aatataattt gttgtcaaa gtaaaaaaaa aaaaaa 1666

```

<210> 513  
 <211> 106  
 <212> PRT  
 <213> Homo sapiens

```

<400> 513
Met Ala His Ala Thr Leu Ser Ala Ala Pro Ser Asn Pro Arg Leu Leu
1      5      10      15
Arg Val Ala Leu Leu Leu Leu Leu Val Gly Ser Arg Arg Ala Ala
20     25     30
Gly Ala Ser Val Val Thr Glu Leu Arg Cys Gln Cys Leu Gln Thr Leu
35     40     45
Gln Gly Ile His Leu Lys Asn Ile Gln Ser Val Asn Val Arg Ser Pro
50     55     60
Gly Pro His Cys Ala Gln Thr Glu Val Ile Ala Thr Leu Lys Asn Gly
65     70     75     80
Lys Lys Ala Cys Leu Asn Pro Ala Ser Pro Met Val Gln Lys Ile Ile
85     90     95
Glu Lys Ile Leu Asn Lys Gly Ser Thr Asn
100    105

```

<210> 514  
 <211> 1064  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <222> (27)...(27)  
 <223> N=A, T, G, or C

```

<400> 514
cacagccggg tgcaggcac ctcccngcc agctctcccg cattctgcac agcttcccg 60
cgcgctctgt gagcccatg gccacgcca cgctctccgc cgccccagc aatccccggc 120
tcctgcgggt ggcgctgctg ctctgctcc tggtgggcag ccggcgcgca gcaggagcgt 180

```

```

ccgtggtcac tgaactgcgc tgccagtgtc tgcagacact gcagggaatt cacctcaaga 240
acatccaaag tgtgaatgta aggtcccccg gacccactg cgcccaaacc gaagtcatag 300
ccacactcaa gaattgggaag aaagcttgct tcaaccccg atcccccag gttcagaaaa 360
tcatcgaaaa gatactgaac aaggggagca ccaactgaca ggagagaagt aagaagctta 420
tcagcgtatc attgacactt cctgcagggt ggtccctgcc cttaccagag ctgaaaatga 480
aaaagagaac agcagctttc tagggacagc tggaaaggga cttaatgtgt ttgactatnt 540
cttacgaggg ttctacttat ttatgtatnt atttttgaaa gcttgtatnt taatattnta 600
catgctgtta tttaaagatg tgagtgtggt tcatcaaaca tagctcagtc ctgattatnt 660
aattggaata tgatgggttt taaatgtgtc attaaactaa tatttagtgg gagaccataa 720
tgtgtcagcc accttgataa atgacagggt ggggaactgg agggtnnggg gattgaaatg 780
caagcaatta gtggatcact gttagggtta gggaaatgat gtacacatct attttttata 840
cttttttttt taaaaaagaa tgtcagttgt tattttattca aattatctca cattatgtgt 900
tcaacattnt tatgctgaag ttcccttag acattttatg tcttgcttgt agggcataat 960
gccttgttta atgtccattc tgcagcgttt ctctttccct tggaaaagag aatttatcat 1020
tactgttaca tttgtacaaa tgacatgata ataaaagtnt tatg 1064

```

&lt;210&gt; 515

&lt;211&gt; 99

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 515

```

Met Lys Val Ser Ala Ala Leu Leu Cys Leu Leu Leu Ile Ala Ala Thr
  1             5             10             15
Phe Ile Pro Gln Gly Leu Ala Gln Pro Asp Ala Ile Asn Ala Pro Val
  20             25             30
Thr Cys Cys Tyr Asn Phe Thr Asn Arg Lys Ile Ser Val Gln Arg Leu
  35             40             45
Ala Ser Tyr Arg Arg Ile Thr Ser Ser Lys Cys Pro Lys Glu Ala Val
  50             55             60
Ile Phe Lys Thr Ile Val Ala Lys Glu Ile Cys Ala Asp Pro Lys Gln
  65             70             75             80
Lys Trp Val Gln Asp Ser Met Asp His Leu Asp Lys Gln Thr Gln Thr
  85             90             95
Pro Lys Thr

```

&lt;210&gt; 516

&lt;211&gt; 757

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 516

```

ggaaccgaga ggctgagact aaccagaaa catccaattc tcaaactgaa gctcgcactc 60
tcgctccag catgaaagtc tctgccgcc ttctgtgcct gctgctcata gcagccacct 120
tcattcccca agggctcgtc cagccagatg caatcaatgc ccagtcacc tgctgttata 180
acttcacca taggaagatc tcagtgcaga ggctcgcgag ctatagaaga atcaccagca 240
gcaagtgtcc caaagaagct gtgatcttca agaccattgt ggccaaggag atctgtgctg 300
acccaagca gaagtgggtt caggattcca tggaccacct ggacaagcaa acccaaaactc 360
cgaagacttg aacactcact ccacaaccca agaattctgca gctaacttat ttcccttag 420
ctttcccccag acaccctggt ttattttatt ataataaatt ttgtttgttg atgtgaaaca 480
ttatgcctta agtaatgtta attcttattt aagttattga tgttttaagt ttatctttca 540
tggtactagt gttttttaga tacagagact tggggaaatt gcttttcctc ttgaaccaca 600
gttctacccc tgggatgttt tgagggtcct tgcaagaatc attaatacaa agaatttttt 660
ttaacattcc aatgcattgc taaaatatta ttgtggaaat gaatattttg taactattac 720
accaataaaa tatatttttg tacaaaaaaa aaaaaaa 757

```

&lt;210&gt; 517

&lt;211&gt; 415

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 517

```

Met Glu Leu Arg Lys Tyr Gly Pro Gly Arg Leu Ala Gly Thr Val Ile
 1           5           10           15
Gly Gly Ala Ala Gln Ser Lys Ser Gln Thr Lys Ser Asp Ser Ile Thr
          20           25           30
Lys Glu Phe Leu Pro Gly Leu Tyr Thr Ala Pro Ser Ser Pro Phe Pro
   35           40           45
Pro Ser Gln Val Ser Asp His Gln Val Leu Asn Asp Ala Glu Val Ala
 50           55           60
Ala Leu Leu Glu Asn Phe Ser Ser Ser Tyr Asp Tyr Gly Glu Asn Glu
 65           70           75           80
Ser Asp Ser Cys Cys Thr Ser Pro Pro Cys Pro Gln Asp Phe Ser Leu
          85           90           95
Ile Asn Phe Asp Arg Ala Phe Leu Pro Ala Leu Tyr Ser Leu Leu Phe
      100           105           110
Leu Leu Gly Leu Leu Gly Asn Gly Ala Val Ala Ala Val Leu Leu Ser
     115           120           125
Arg Arg Thr Ala Leu Ser Ser Thr Asp Thr Phe Leu Leu His Leu Ala
    130           135           140
Val Ala Asp Thr Leu Leu Val Leu Thr Leu Pro Leu Trp Ala Val Asp
   145           150           155           160
Ala Ala Val Gln Trp Val Phe Gly Ser Gly Leu Cys Lys Val Ala Gly
      165           170           175
Ala Leu Phe Asn Ile Asn Phe Tyr Ala Gly Ala Leu Leu Leu Ala Cys
     180           185           190
Ile Ser Phe Asp Arg Tyr Leu Asn Ile Val His Ala Thr Gln Leu Tyr
    195           200           205
Arg Arg Gly Pro Pro Ala Arg Val Thr Leu Thr Cys Leu Ala Val Trp
   210           215           220
Gly Leu Cys Leu Leu Phe Ala Leu Pro Asp Phe Ile Phe Leu Ser Ala
  225           230           235           240
His His Asp Glu Arg Leu Asn Ala Thr His Cys Gln Tyr Asn Phe Pro
      245           250           255
Gln Val Gly Arg Thr Ala Leu Arg Val Leu Gln Leu Val Ala Gly Phe
     260           265           270
Leu Leu Pro Leu Leu Val Met Ala Tyr Cys Tyr Ala His Ile Leu Ala
    275           280           285
Val Leu Leu Val Ser Arg Gly Gln Arg Arg Leu Arg Ala Met Arg Leu
   290           295           300
Val Val Val Val Val Ala Phe Ala Leu Cys Trp Thr Pro Tyr His
  305           310           315           320
Leu Val Val Leu Val Asp Ile Leu Met Asp Leu Gly Ala Leu Ala Arg
      325           330           335
Asn Cys Gly Arg Glu Ser Arg Val Asp Ala Lys Ser Val Thr Ser Gly
    340           345           350
Leu Gly Tyr Met His Cys Cys Leu Asn Pro Leu Leu Tyr Ala Phe Val
   355           360           365
Gly Val Lys Phe Arg Glu Arg Met Trp Met Leu Leu Arg Leu Gly
  370           375           380
Cys Pro Asn Gln Arg Gly Leu Gln Arg Gln Pro Ser Ser Ser Arg Arg
 385           390           395           400
Asp Ser Ser Trp Ser Glu Thr Ser Glu Ala Ser Tyr Ser Gly Leu
      405           410           415

```

&lt;210&gt; 518

&lt;211&gt; 1703

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens



&lt;400&gt; 518

```
atggagttga ggaagtacgg ccctggaaga ctggcgggga cagttatagg aggagctgct 60
cagagtaaat cacagactaa atcagactca atcacaaaag agttcctgcc aggcctttac 120
acagcccctt cctccccgtt cccgccctca caggtagtg accaccaagt gctaaatgac 180
gccgaggttg ccgccctcct ggagaacttc agctcttcct atgactatgg agaaaacgag 240
agtgactcgt gctgtacctc cccgccctgc ccacaggact tcagcctgaa cttcgaccgg 300
gccttcctgc cagccctcta cagcctcctc tttctgctgg ggctgctggg caacggcgcg 360
gtggcagcgg tgctgctgag ccggcggaaca gccctgagca gcaccgacac cttcctgctc 420
cacctagttg tagcagacac gctgctggtg ctgacactgc cgctctgggc agtggagcgt 480
gccgtccagt gggctctttg ctctggcctc tgcaaagtgg cagggtgccct cttcaacatc 540
aacttctacg caggagccct cctgctggcc tgcacagct ttgaccgcta cctgaacata 600
gttcatgcca cccagctcta ccgccggggg ccccccggcc gcgtgaccct cacctgctg 660
gctgtctggg ggctctgctt gcttttcgcc ctcccagact tcatcttctt gtcggccccac 720
cacgacgagc gcctcaacgc caccactgc caatacaact tcccacagggt gggccgcacg 780
gctctgcggg tgctgcagct ggtggctggc tttctgctgc cctgctggt catggcctac 840
tgctatgccc acatcctggc cgtgctgctg gtttccaggg gccagcgggc cctgcgggcc 900
atgcggctgg tgggtggtgg cgtggtggcc tttgccctct gctggacccc ctatcacctg 960
gtggtgctgg tggacatcct catggacctg ggcgctttgg cccgcaactg tggccgagaa 1020
agcagggtag acgtggccaa gtcggtcacc tcaggcctgg gctacatgca ctgctgctc 1080
aaccgctgc tctatgcctt tgtaggggtc aagttccggg agcggatgtg gatgctgctc 1140
ttgcgcctgg gctgccccaa ccagagaggg ctccagaggc agccatcgct tcccgcggg 1200
gattcatcct ggtctgagac ctacagaggc tctactcgg gcttgtagg ccggaatccg 1260
ggctccccct tcgcccacag tctgacttcc ccgcattcca ggctcctccc tccctctgcc 1320
ggctctggct ctcccccaata tctcgtctcc cgggactcac tggcagcccc agcaccacca 1380
ggtctcccggaagccaccc tcccagctct gaggactgca ccattgctgc tccttagctg 1440
ccaagcccca tctgcccgcg cgagggtggct gcctggagcc ccactgccc tctcatttgg 1500
aaactaaaac ttcatcttcc ccaagtgcgg ggagtacaag gcatggcgta gagggtgctg 1560
cccatgaag ccacagccca ggctccagc tcagcagtga ctgtggccat ggtccccaag 1620
acctctatat ttgctctttt atttttatgt ctaaaatcct gcttaaaact tttcaataaa 1680
caagatcgctc aggacaaaaa aaa 1703
```

&lt;210&gt; 519

&lt;211&gt; 60

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Aligent oligonucleotide

&lt;400&gt; 519

```
gtactttgtg tttaatatat ctgggtgatg gatcacaaca catcaataaa ctgacttacc 60
```

&lt;210&gt; 520

&lt;211&gt; 60

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic sequence

&lt;400&gt; 520

```
atttacttgc atatgtaaac cattgctgtg ccattcaatg tttgatgcat aattggacct 60
```

&lt;210&gt; 521

&lt;211&gt; 60

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic sequence

&lt;400&gt; 521

agatggagga agcatctgag tttagacca tggctgttac agggatcatg taaacttgct 60

&lt;210&gt; 522

&lt;211&gt; 60

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic sequence

&lt;400&gt; 522

tggagattta aacactgagg ttctgttca aactgtgagt tctgttcttt gtgagaaatt 60

&lt;210&gt; 523

&lt;211&gt; 60

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic sequence

&lt;400&gt; 523

tgtacctgaa aaatgggaag agtcgggacc acagtttatt accaattctg aggaagtccg 60

&lt;210&gt; 524

&lt;211&gt; 60

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic sequence

&lt;400&gt; 524

atccatttac tcagctggag aggagacatc aagaacatgc cagacacatt tcttttggtg 60

&lt;210&gt; 525

&lt;211&gt; 60

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic sequence

&lt;400&gt; 525

ctggatgttg aattgccacc tgtttgctgt gacatagata tttaaatttc ttagtgcttc 60

&lt;210&gt; 526

&lt;211&gt; 60

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> synthetic sequence

<400> 526

ctggaaacct ctggagggtca tctcggctgt tcctgagaaa taaaaagcct gtcatttcaa 60

<210> 527

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic sequence

<400> 527

tgttgacgtc tcctttaagc attctgttga ccctggatgt tgaattgccca cctgtttgct 60

<210> 528

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic sequence

<400> 528

actattttctt cagagtttgt catatactgc ttgtcatctg catgtctact cagcatttga 60

<210> 529

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic sequence

<400> 529

agactgttaa aacaaaaagt tctgttcctg cctcagatga cgcctatcca gaaatagaaa 60

<210> 530

<211> 1863

<212> DNA

<213> Homo sapiens

<400> 530

gtgcagcctg	atgcgcaacg	tggggactca	ggcgcgctgg	gcggcaggag	ttgcttccgg	60
ccgtgttggg	ggtctgaatt	gagaagccgc	gactaaggga	agatggagac	aatactggag	120
cagcagcggc	gctatcatga	ggagaaggaa	cggtcatggg	acgtcatggc	taaagagatg	180
ctcaccaaga	agtccacgct	ccgggaccag	atcaattctg	atcaccgcac	tcgggccatg	240
caagataggt	atatggaggt	cagtgggaac	ctgagggatt	tgtatgatga	taaggatgga	300
ttacgaaagg	aggagctcaa	tgccatttca	ggacccaatg	agtttgctga	attctataat	360
agactcaagc	aaataaagga	attccaccgg	aagcacccaa	atgagatctg	tgtgccaatg	420
tcagtggaa	ttgaggaact	cctgaaggct	cgagagaatc	caagtgaaga	ggcacaaaac	480
ttggtggagt	tcacagatga	ggagggatat	ggtcgttatc	tcgatctcca	tgactgttac	540
ctcaagtaca	ttaacctgaa	ggcatctgag	aagctggatt	atatcacata	cctgtccatc	600
tttgaccaat	tatttgacat	tcctaaagaa	aggaagaatg	cagagtataa	gagataccta	660
gagatgctgc	ttgagtacct	tcaggattac	acagatagag	tgaagcctct	ccaagatcag	720
aatgaacttt	ttgggaagat	tcaggctgag	tttgagaaga	aatgggagaa	tgggaccttt	780

```

cctggatggc cgaaagagac aagcagtgcc ctgacccatg ctggagccca tcttgacctc 840
tctgcattct cctcctggga ggagttggct tctctgggtt tggacagatt gaaatctgct 900
ctcttagctt taggcttgaa atgtggcggg accctagaag agcgagccca gagactattc 960
agtaccaaag gaaagtcctt ggagtcactt gatacctctt tgtttgccaa aaatcccaag 1020
tcaaagggca ccaagcgaga cactgaaagg aacaaagaca ttgcttttct agaagcccag 1080
atctatgaat atgtagagat tctcggggaa cagcgacatc tcactcatga aaatgtacag 1140
cgcaagcaag ccaggacagg agaagagcga gaagaagagg aagaagagca gatcagtgag 1200
agtgagagtg aagatgaaga gaacgagatc atttacaacc ccaaaaacct gccacttggc 1260
tgggatggca aacctattcc ctactggctg tataagcttc atggcctaaa tatcaactac 1320
aactgtgaga tttgtggaaa ctacacctac cgagggccca aagccttcca gcgacacttt 1380
gctgaatggc gtcattgctca tggcatgagg tgtttgggca tcccaaacac tgctcacttt 1440
gctaattgta cacagattga agatgctgtc tccttgtggg ccaaactgaa attgcagaag 1500
gcttcagaac gatggcagcc tgacactgag gaagaatatg aagactcaag tgggaattgt 1560
gtgaataaga agacatacga ggatctgaaa agacaaggac tgctctagtg ttgagggatg 1620
tagctcagct tttgggctag cccaggcttc cctaagatct gctttttcta tttctcccaa 1680
ccaaatcctc ttaaagaccc tttgctatgt agtctcatgg tctagcatgc atcttgtaga 1740
aacaaggcat gctggcagat tgcaggggtt agatgtgttt tatctgtttt atattttaaa 1800
agattctgcc agaaaataaa accagacctt gttctaaaaa aaaaaaaaaa aaaaaaaaaa 1860
aaa

```

1863

<210> 531  
 <211> 501  
 <212> PRT  
 <213> Homo sapiens

<400> 531

```

Met Glu Thr Ile Leu Glu Gln Gln Arg Arg Tyr His Glu Glu Lys Glu
1      5      10      15
Arg Leu Met Asp Val Met Ala Lys Glu Met Leu Thr Lys Lys Ser Thr
20     25     30
Leu Arg Asp Gln Ile Asn Ser Asp His Arg Thr Arg Ala Met Gln Asp
35     40     45
Arg Tyr Met Glu Val Ser Gly Asn Leu Arg Asp Leu Tyr Asp Asp Lys
50     55     60
Asp Gly Leu Arg Lys Glu Glu Leu Asn Ala Ile Ser Gly Pro Asn Glu
65     70     75     80
Phe Ala Glu Phe Tyr Asn Arg Leu Lys Gln Ile Lys Glu Phe His Arg
85     90     95
Lys His Pro Asn Glu Ile Cys Val Pro Met Ser Val Glu Phe Glu Glu
100    105    110
Leu Leu Lys Ala Arg Glu Asn Pro Ser Glu Glu Ala Gln Asn Leu Val
115    120    125
Glu Phe Thr Asp Glu Glu Gly Tyr Gly Arg Tyr Leu Asp Leu His Asp
130    135    140
Cys Tyr Leu Lys Tyr Ile Asn Leu Lys Ala Ser Glu Lys Leu Asp Tyr
145    150    155    160
Ile Thr Tyr Leu Ser Ile Phe Asp Gln Leu Phe Asp Ile Pro Lys Glu
165    170    175
Arg Lys Asn Ala Glu Tyr Lys Arg Tyr Leu Glu Met Leu Leu Glu Tyr
180    185    190
Leu Gln Asp Tyr Thr Asp Arg Val Lys Pro Leu Gln Asp Gln Asn Glu
195    200    205
Leu Phe Gly Lys Ile Gln Ala Glu Phe Glu Lys Lys Trp Glu Asn Gly
210    215    220
Thr Phe Pro Gly Trp Pro Lys Glu Thr Ser Ser Ala Leu Thr His Ala
225    230    235    240
Gly Ala His Leu Asp Leu Ser Ala Phe Ser Ser Trp Glu Glu Leu Ala
245    250    255
Ser Leu Gly Leu Asp Arg Leu Lys Ser Ala Leu Leu Ala Leu Gly Leu
260    265    270
Lys Cys Gly Gly Thr Leu Glu Glu Arg Ala Gln Arg Leu Phe Ser Thr

```

275 280 285  
 Lys Gly Lys Ser Leu Glu Ser Leu Asp Thr Ser Leu Phe Ala Lys Asn  
 290 295 300  
 Pro Lys Ser Lys Gly Thr Lys Arg Asp Thr Glu Arg Asn Lys Asp Ile  
 305 310 315 320  
 Ala Phe Leu Glu Ala Gln Ile Tyr Glu Tyr Val Glu Ile Leu Gly Glu  
 325 330 335  
 Gln Arg His Leu Thr His Glu Asn Val Gln Arg Lys Gln Ala Arg Thr  
 340 345 350  
 Gly Glu Glu Arg Glu Glu Glu Glu Glu Glu Gln Ile Ser Glu Ser Glu  
 355 360 365  
 Ser Glu Asp Glu Glu Asn Glu Ile Ile Tyr Asn Pro Lys Asn Leu Pro  
 370 375 380  
 Leu Gly Trp Asp Gly Lys Pro Ile Pro Tyr Trp Leu Tyr Lys Leu His  
 385 390 395 400  
 Gly Leu Asn Ile Asn Tyr Asn Cys Glu Ile Cys Gly Asn Tyr Thr Tyr  
 405 410 415  
 Arg Gly Pro Lys Ala Phe Gln Arg His Phe Ala Glu Trp Arg His Ala  
 420 425 430  
 His Gly Met Arg Cys Leu Gly Ile Pro Asn Thr Ala His Phe Ala Asn  
 435 440 445  
 Val Thr Gln Ile Glu Asp Ala Val Ser Leu Trp Ala Lys Leu Lys Leu  
 450 455 460  
 Gln Lys Ala Ser Glu Arg Trp Gln Pro Asp Thr Glu Glu Glu Tyr Glu  
 465 470 475 480  
 Asp Ser Ser Gly Asn Val Val Asn Lys Lys Thr Tyr Glu Asp Leu Lys  
 485 490 495  
 Arg Gln Gly Leu Leu  
 500

<210> 532  
 <211> 3326  
 <212> DNA  
 <213> Homo sapiens

<400> 532  
 ctgagcctcc cgtgccggcc ggccggccgg gggaacaggc gggcgctcgg gggcgctcgg 60  
 gggggcgggg ggagttccgg ttccggttct ttgtgcggct gcatcgccgg ctccgggaag 120  
 atggcgggcc gggcggttt ccagtcctgt gctccaagcg gcggcgccgg agcctcagga 180  
 gggcggggcg cggctgctgc cttgggcccgg ggcggaactc cggggcctcc tgtgcgaatg 240  
 ggcccggtc cgggtcaagg gctgtaccgc tccccgatgc ccggagcggc ctatccgaga 300  
 ccaggtatgt tgccaggcag ccgaatgaca cctcagggac cttccatggg accccctggc 360  
 tatgggggga acccttcagt ccgacctggc ctggcccagt cagggatgga tcagtccgcg 420  
 aagagacctg cccctcagca gatccagcag gtccagcagc aggcggtcca aaatcgaaac 480  
 cacaatgcaa agaaaaagaa gatggctgac aaaattctac ctcaaaggat tcgtgaactg 540  
 gtaccagaat cccaggccta tatggatctc ttggcttttg aaaggaaact ggaccagact 600  
 atcatgagga aacggctaga tatccaagag gccttgaaac gtcccatcaa gcaaaaacgg 660  
 aagctgcgaa ttttcatttc taacactttc aatccggcta agtcagatgc cgaggatggg 720  
 gaaggtagcg tggcttcctg ggagcttcgg gtagaaggac ggctcctgga ggattcagcc 780  
 ttgtccaaat atgatgccac taaacaaaag aggaagttct cttccttttt taagtccttg 840  
 gtgattgaac tggacaaaga cctgtatggg ccagacaacc atctggtaga atggcacagg 900  
 accgccacta cccaggagac cgatggcttt caggatgaagc ggccgggaga cgtgaatgta 960  
 cgggtgtactg tctactgat gctggattac cagcctcccc agtttaaatt agacccccgc 1020  
 ctagctcgac tcttgggcat ccatacccag actcgtccag tgatcatcca agcactgttg 1080  
 caatatatta agacacataa gctccaggac cctcagcagc gggagtttgt catctgtgac 1140  
 aagtacctgc agcacatctt tgagtctcaa cgttaagat tttcagagat cctcagcgg 1200  
 ctccatgcct tgcttatgcc accagaacct atcatcatta atcatgtcat cagtgttgac 1260  
 ccgaatgatc agaaaaagac agcttgttat gacattgatg ttgaagtgga tgacaccttg 1320  
 aagaccaga tgaattcttt tctgctgtcc actgccagcc aacaggagat tgctactcta 1380  
 gacaacaaga tccatgagac aatagaaacc atcaaccagg tgaagactca gcgggagttc 1440

```

atgctgagct ttgccagaga ccctcagggt ttcataatg actggcttca gtcccagtgc 1500
agggacctca agacaatgac tgatgtggtg ggtaaccagc aggaggagcg ccgagctgag 1560
ttctacttcc agccctgggc tcaggaggct gtgtgccgat acttctactc caaggtgcag 1620
cagagacgac aagaattaga gcaagccctg ggaatccgga atacataggg cctctccac 1680
agccctgatt cgactgcacc aattcttgat ttgggcctg tgctgcctgc ctcatagtat 1740
ctgccttggt cttgcttggg gcgttccagg ggatgctgtt ggttcaagga caacaccaga 1800
atgaagaggg tctcacaaga cacctgttat cctctctttt caccctatct cttccacacc 1860
ccagcttccc tttgcccac aaagtccca tgtgcctgta cctcccctg gtctacatag 1920
gacctctaga tagtgtaga gagagaacat gtagtggtaa tgagtgcctg gaatggattg 1980
ggcctcaggc caggtggtct tcaaggggac cagctaactg atcctgccct tcagagaccc 2040
aggagttggg agcttctgct ccttctccaa gactcaggcc tgtgggcact ctataagcta 2100
gttgatcttg gctctcctga taacagaatc caatttctt ccttccctcc acaggttttg 2160
aacaaactct ccttctactt gttgccctgt agcactacag aaaccctggt tcttgggctg 2220
cactgagccc caggtcagtc cccagccctc tgggttggcc tgctgtcagt gcttctctca 2280
ctccttagtt ggggtccaca tcagtattgg agttttgtt tttattgctc cctcccagac 2340
actccctgtg gctgcccttt gtgattccct cagatctgcc ctaatcccgg gcatttgggt 2400
gggggaatct tgccttcccc tttcagagcc ccagggatct catctgggga actgtcattg 2460
ccagcagagg ctgttccctc ctgctgtttg gagatgtgac tcattcattc actcactcca 2520
cctgctctct gcatccctta atggagaaac gggcctaaaa ccaaacgggt aaaaagccct 2580
gggccatccc tgtcttctct tcccttgtct gccagttga cacctactgg tgacttctag 2640
ggcactgagg agtgaaagcg cctagggctg gagaatagcg ctgagttggg tttgtgactc 2700
ttccctctcc ctgcctcaca ggattgtgac tcccagccc ctgcccctca agcttcagac 2760
cctcaggtga gcagcaggac cttgtgatct tggccccttg gatctgagat ggtttttgca 2820
tctttccagg agagcctcac attcttcttc caggttgtat caccctcgag ttagcatatc 2880
ccaggctcgc agactcaaca cagcaagggt gggagacagc tgggcacaaa gggggaattc 2940
cgttcagcat gggctctaaa cccacagaac tgacaaagcc cctgcttccc caccctctcc 3000
tcaggctcct gcgagcacac cccaccccc aaatccctcc ctgttctaca ctggggcagc 3060
cagaattttc tcccgtctt ccccttctct ccatttctcc tcccttgaaa ggttgacact 3120
ggacaacctt ggggcagctg agccctggcc gcctcctggc tggaacctg agaaggaagc 3180
tcagtacttc ccacagtgtc cctgttgata actgttttta ttaactgaat tgttttttcc 3240
atggaccaaa cttttttttg tactgtcccc ttattgatgt taccagttt taataaaaga 3300
atcttctgaa aaaaaaaaa aaaaaa 3326

```

<210> 533  
 <211> 515  
 <212> PRT  
 <213> Homo sapiens

<400> 533

```

Met Ala Ala Arg Ala Gly Phe Gln Ser Val Ala Pro Ser Gly Gly Ala
 1          5          10          15
Gly Ala Ser Gly Gly Ala Gly Ala Ala Ala Ala Leu Gly Pro Gly Gly
 20          25          30
Thr Pro Gly Pro Pro Val Arg Met Gly Pro Ala Pro Gly Gln Gly Leu
 35          40          45
Tyr Arg Ser Pro Met Pro Gly Ala Ala Tyr Pro Arg Pro Gly Met Leu
 50          55          60
Pro Gly Ser Arg Met Thr Pro Gln Gly Pro Ser Met Gly Pro Pro Gly
 65          70          75          80
Tyr Gly Gly Asn Pro Ser Val Arg Pro Gly Leu Ala Gln Ser Gly Met
 85          90          95
Asp Gln Ser Arg Lys Arg Pro Ala Pro Gln Gln Ile Gln Gln Val Gln
100          105          110
Gln Gln Ala Val Gln Asn Arg Asn His Asn Ala Lys Lys Lys Lys Met
115          120          125
Ala Asp Lys Ile Leu Pro Gln Arg Ile Arg Glu Leu Val Pro Glu Ser
130          135          140
Gln Ala Tyr Met Asp Leu Leu Ala Phe Glu Arg Lys Leu Asp Gln Thr
145          150          155          160
Ile Met Arg Lys Arg Leu Asp Ile Gln Glu Ala Leu Lys Arg Pro Ile
165          170          175

```

Lys Gln Lys Arg Lys Leu Arg Ile Phe Ile Ser Asn Thr Phe Asn Pro  
 180 185 190  
 Ala Lys Ser Asp Ala Glu Asp Gly Glu Gly Thr Val Ala Ser Trp Glu  
 195 200 205  
 Leu Arg Val Glu Gly Arg Leu Leu Glu Asp Ser Ala Leu Ser Lys Tyr  
 210 215 220  
 Asp Ala Thr Lys Gln Lys Arg Lys Phe Ser Ser Phe Phe Lys Ser Leu  
 225 230 235 240  
 Val Ile Glu Leu Asp Lys Asp Leu Tyr Gly Pro Asp Asn His Leu Val  
 245 250 255  
 Glu Trp His Arg Thr Ala Thr Thr Gln Glu Thr Asp Gly Phe Gln Val  
 260 265 270  
 Lys Arg Pro Gly Asp Val Asn Val Arg Cys Thr Val Leu Leu Met Leu  
 275 280 285  
 Asp Tyr Gln Pro Pro Gln Phe Lys Leu Asp Pro Arg Leu Ala Arg Leu  
 290 295 300  
 Leu Gly Ile His Thr Gln Thr Arg Pro Val Ile Ile Gln Ala Leu Trp  
 305 310 315 320  
 Gln Tyr Ile Lys Thr His Lys Leu Gln Asp Pro His Glu Arg Glu Phe  
 325 330 335  
 Val Ile Cys Asp Lys Tyr Leu Gln Gln Ile Phe Glu Ser Gln Arg Met  
 340 345 350  
 Lys Phe Ser Glu Ile Pro Gln Arg Leu His Ala Leu Leu Met Pro Pro  
 355 360 365  
 Glu Pro Ile Ile Ile Asn His Val Ile Ser Val Asp Pro Asn Asp Gln  
 370 375 380  
 Lys Lys Thr Ala Cys Tyr Asp Ile Asp Val Glu Val Asp Asp Thr Leu  
 385 390 395 400  
 Lys Thr Gln Met Asn Ser Phe Leu Leu Ser Thr Ala Ser Gln Gln Glu  
 405 410 415  
 Ile Ala Thr Leu Asp Asn Lys Ile His Glu Thr Ile Glu Thr Ile Asn  
 420 425 430  
 Gln Leu Lys Thr Gln Arg Glu Phe Met Leu Ser Phe Ala Arg Asp Pro  
 435 440 445  
 Gln Gly Phe Ile Asn Asp Trp Leu Gln Ser Gln Cys Arg Asp Leu Lys  
 450 455 460  
 Thr Met Thr Asp Val Val Gly Asn Pro Glu Glu Glu Arg Arg Ala Glu  
 465 470 475 480  
 Phe Tyr Phe Gln Pro Trp Ala Gln Glu Ala Val Cys Arg Tyr Phe Tyr  
 485 490 495  
 Ser Lys Val Gln Gln Arg Arg Gln Glu Leu Glu Gln Ala Leu Gly Ile  
 500 505 510  
 Arg Asn Thr  
 515

&lt;210&gt; 534

&lt;211&gt; 1662

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 534

gttaaagccg gtcggaactg ctccggaggg cacgggctcc gtaggcaçca actgcaagga 60  
 cccctccccc tgcgggcgct cccatggcac agttcgcggt cgagagtgc ctgcactcgc 120  
 tgcttcagct ggatgcaccc atccccaatg caccctcgc gcgctggcag cgcaaagcca 180  
 aggaagccgc aggcgccggc cctcaccca tgcgggccgc caaccgatcc cacagcgccg 240  
 gcaggactcc gggccgaact cctggcaa atccagttccaa gggttcagacc actcctagca 300  
 aacctggcgg tgaccgctat atccccatc gcagtgctgc ccagatggag gtggccagct 360  
 tcctcctgag caaggagaac cagcctgaaa acagccagac gccaccaag aaggacatc 420  
 agaaagcctg ggctttgaac ctgaacggtt ttgatgtaga ggaagccaag atccttcggc 480  
 tcagtggaaa accacaaaat gcgccagagg gttatcagaa cagactgaaa gtactctaca 540

```

gccaaaaggc cactcctggc tccagccgga agacctgccg ttacattcct tccctgccag 600
accgtatcct ggatgcgcct gaaatccgaa atgactatta cctgaacctt gtggattgga 660
gttctgggaa tgtactggcc gtggcactgg acaacagtgt gtacctgtgg agtgcaagct 720
ctggtgacat cctgcagctt ttgcaaatgg agcagcctgg ggaatatata tcctctgtgg 780
cctggatcaa agagggaac tacttggtcg tgggcaccag cagtgtctgag gtgcagctat 840
gggatgtgca gcagcagaaa cggcttcgaa atatgaccag tcactctgcc cgagtgggct 900
ccctaagctg gaacagctat atcctgtcca gtggttcacg ttctggccac atccaccacc 960
atgatgttcg ggtagcagaa caccatgtgg ccacactgag tggccacagc caggaagtgt 1020
gtgggctgcg ctgggccccg gatggacgac atttggccag tggtgtaaat gataacttgg 1080
tcaatgtgtg gcctagtgtt cctggagagg gtggctgggt tcctctgcag acattcaccc 1140
agcatcaagg ggctgtcaag gccgtagcat ggtgtccctg gcagtccaat gtcctggcaa 1200
caggaggggg caccagtgtat cgacacattc gcatctggaa tgtgtgctct ggggctgtc 1260
tgagtgcctg ggatgcccac tcccaggtgt gctccatcct ctgggtctcc cattacaagg 1320
agctcatctc aggccatggc tttgcacaga accagctagt tatttggag taccacaacca 1380
tggccaaggt ggctgaactc aaaggtcaca catcccggt cctgagctctg accatgagcc 1440
cagatggggc cacagtggca tccgcagcag cagatgagac cctgaggcta tggcgctgtt 1500
ttgagttgga cctgcgcgg cggcgggagc gggagaaggc cagtgcagcc aaaagcagcc 1560
tcatccacca aggcattcgc tgaagaccaa cccatcacct cagttgtttt ttatttttct 1620
aataaagtca tgtctccctt caaaaaaaaa aaaaaaaaaa aa 1662

```

&lt;210&gt; 535

&lt;211&gt; 499

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 535

```

Met Ala Gln Phe Ala Phe Glu Ser Asp Leu His Ser Leu Leu Gln Leu
1      5      10      15
Asp Ala Pro Ile Pro Asn Ala Pro Pro Ala Arg Trp Gln Arg Lys Ala
20      25      30
Lys Glu Ala Ala Gly Pro Ala Pro Ser Pro Met Arg Ala Ala Asn Arg
35      40      45
Ser His Ser Ala Gly Arg Thr Pro Gly Arg Thr Pro Gly Lys Ser Ser
50      55      60
Ser Lys Val Gln Thr Thr Pro Ser Lys Pro Gly Gly Asp Arg Tyr Ile
65      70      75      80
Pro His Arg Ser Ala Ala Gln Met Glu Val Ala Ser Phe Leu Leu Ser
85      90      95
Lys Glu Asn Gln Pro Glu Asn Ser Gln Thr Pro Thr Lys Lys Glu His
100     105     110
Gln Lys Ala Trp Ala Leu Asn Leu Asn Gly Phe Asp Val Glu Glu Ala
115     120     125
Lys Ile Leu Arg Leu Ser Gly Lys Pro Gln Asn Ala Pro Glu Gly Tyr
130     135     140
Gln Asn Arg Leu Lys Val Leu Tyr Ser Gln Lys Ala Thr Pro Gly Ser
145     150     155     160
Ser Arg Lys Thr Cys Arg Tyr Ile Pro Ser Leu Pro Asp Arg Ile Leu
165     170     175
Asp Ala Pro Glu Ile Arg Asn Asp Tyr Tyr Leu Asn Leu Val Asp Trp
180     185     190
Ser Ser Gly Asn Val Leu Ala Val Ala Leu Asp Asn Ser Val Tyr Leu
195     200     205
Trp Ser Ala Ser Ser Gly Asp Ile Leu Gln Leu Leu Gln Met Glu Gln
210     215     220
Pro Gly Glu Tyr Ile Ser Ser Val Ala Trp Ile Lys Glu Gly Asn Tyr
225     230     235     240
Leu Ala Val Gly Thr Ser Ser Ala Glu Val Gln Leu Trp Asp Val Gln
245     250     255
Gln Gln Lys Arg Leu Arg Asn Met Thr Ser His Ser Ala Arg Val Gly
260     265     270
Ser Leu Ser Trp Asn Ser Tyr Ile Leu Ser Ser Gly Ser Arg Ser Gly

```



275 280 285  
 His Ile His His His Asp Val Arg Val Ala Glu His His Val Ala Thr  
 290 295 300  
 Leu Ser Gly His Ser Gln Glu Val Cys Gly Leu Arg Trp Ala Pro Asp  
 305 310 315 320  
 Gly Arg His Leu Ala Ser Gly Gly Asn Asp Asn Leu Val Asn Val Trp  
 325 330 335  
 Pro Ser Ala Pro Gly Glu Gly Gly Trp Val Pro Leu Gln Thr Phe Thr  
 340 345 350  
 Gln His Gln Gly Ala Val Lys Ala Val Ala Trp Cys Pro Trp Gln Ser  
 355 360 365  
 Asn Val Leu Ala Thr Gly Gly Gly Thr Ser Asp Arg His Ile Arg Ile  
 370 375 380  
 Trp Asn Val Cys Ser Gly Ala Cys Leu Ser Ala Val Asp Ala His Ser  
 385 390 395 400  
 Gln Val Cys Ser Ile Leu Trp Ser Pro His Tyr Lys Glu Leu Ile Ser  
 405 410 415  
 Gly His Gly Phe Ala Gln Asn Gln Leu Val Ile Trp Lys Tyr Pro Thr  
 420 425 430  
 Met Ala Lys Val Ala Glu Leu Lys Gly His Thr Ser Arg Val Leu Ser  
 435 440 445  
 Leu Thr Met Ser Pro Asp Gly Ala Thr Val Ala Ser Ala Ala Ala Asp  
 450 455 460  
 Glu Thr Leu Arg Leu Trp Arg Cys Phe Glu Leu Asp Pro Ala Arg Arg  
 465 470 475 480  
 Arg Glu Arg Glu Lys Ala Ser Ala Ala Lys Ser Ser Leu Ile His Gln  
 485 490 495  
 Gly Ile Arg

**THIS PAGE BLANK (USPTO)**

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**